JPII 13 (3) (2024) 459-470

Jurnal Pendidikan IPA Indonesia

https://journal.unnes.ac.id/journals/jpii

VALIDATION AND IMPLEMENTATION OF 3-DIMENSIONAL SCIENTIFIC LITERACY TEST (LISA3D TEST): MEASURING SCIENTIFIC LITERACY FOR SENIOR HIGH SCHOOL STUDENTS BASED ON SCIENTIFIC REASONING, SCIENTIFIC INQUIRY, AND NATURE OF SCIENCE

P. Susongko[1*], N. B. A. Wahab[2], Y. Arfiani[3], M. Kusuma[4]

[1,3,4]Universitas Pancasakti Tegal, Indonesia
[2]Universitas Pendidikan Sultan Idris, Malaysia

DOI: 10.15294/jpii.v13i3.6309

Accepted: June 15th, 2024. Approved: August 29th, 2024. Published: August 30th 2024

ABSTRACT

This research aims to validate and implement an instrument to measure the scientific literacy of high school science program students, including the abilities of Scientific Reasoning, Scientific Inquiry, and Nature of Science (NOS), called the Lisa3D Test. These three instruments were validated using the Rasch model approach and presented as a digital application. This research method was a descriptive survey with a quantitative approach. The research started from the test design, validation, and implementation stages. The types of validity used included content, psychometric, construct, and external validity. For construct validity, the instrument was tested on 197 high school students, while the implementation stage involved 407 students. The research results show that the Lisa3D Test has fulfilled all aspects of validity. The Lisa3D test has been successfully implemented via the page www.lisatest.id with a cut of 56% for the Scientific Reasoning test, 50% for the Scientific Inquiry test, and 56.25% for the NOS test. This research produces a standard instrument to measure the scientific literacy of high school science program students based on three dimensions, namely Scientific Reasoning, Scientific Inquiry, and Nature of Science (NOS). The Lisa3D test can be implemented on a broader scale and can be an additional criterion for graduating high school students so that all graduates can master science comprehensively.

Keywords: inquiry, Lisa3D test, nature of science, reasoning, scientific, validation

INTRODUCTION

According to the policy of the Ministry of Education and Culture, Research and Technology of the Republic of Indonesia, the minimum literacy competency assessment is not only the ability to read but also the ability to analyze reading, the ability to understand or comprehend the concepts behind a piece of writing. It is carried out not based on subjects but on mastery of content or material. Susongko (2019-2021) conducted initial research related to literacy assessment based on the ability to analyze a reading or case (Susongko, 2019; 2020; 2021). This research measured the scientific literacy skills of senior high school students in the natural science program using the PISA 2015 standards with the Rasch model application approach. This research produced an integrated science-based scientific literacy test instrument (SLiSIS Test) that meets all aspects of Messick's validity (Rawand & Firoozi, 2016; Anselmi et al., 2015). The SLiSIS Test as a minimum test of scientific literacy has fulfilled the main objective of analyzing readings or cases; however, it needs to be developed further as a standard for fulfilling scientific literacy at the high school level. This is because high school competency standards are pretty comprehensi-

*Correspondence Address
E-mail: purwosusongko@upstegal.ac.id

ve, including the ability to apply knowledge and technology logically, critically, creatively, and innovatively, as well as the more complex demands for achieving scientific literacy at the high school level. Based on this, it is necessary to develop a scientific literacy test that is more comprehensive and meets standards for high school.

Shamos (1995) divides the dimensions of scientific literacy into three levels: cultural scientific literacy, functional scientific literacy, and valid scientific literacy. The dimensions of true scientific literacy contain the mental qualities that John Dewey called the habit of scientific thinking, and almost a century ago, he proposed as the main reason for the obligation of science education (Hanson, 2016; Fives et al., 2014; Osborne, 2023). Furthermore, Hanson (2016) explains that someone classified as having true scientific literacy can understand the scientific process of developing knowledge, understand the importance of observation and experimentation in science, and have the ability to question. Other skills are using logic for induction and deduction, relying on evidence, having a correct understanding of the nature of science, and having a basic understanding of science's history, values , and assumptions (Hanson, 2016). Based on these opinions, it can be concluded that actual scientific literacy skill is not only knowledge of the basic facts established by science, but this concept includes more than that. There are at least three aspects of scientific literacy: comprehending content, understanding scientific practice, and understanding science as a social process.

Based on more than 15 research results on the nature of science, Mc Comas and Nouri (2016) have formulated indicators for how someone understands the nature of science, history, values, and assumptions about science. These indicators include a person's understanding of science that scientific knowledge is not entirely objective, that scientists use creativity, that scientific knowledge is tentative but long-lasting, and that scientific knowledge is socially and culturally embedded. Another understanding of science is that law and theory are different types of knowledge. Scientific knowledge is based on empirical research; there is no universal, gradual scientific method and a difference between observation and conclusion. Likewise, it must be understood that science cannot answer all questions; cooperation and collaboration are part of the development of science, and there are differences between science and technology, as well as the importance of experiments in the development of science (Erduran et al., 2014; McComas, 2013). When referring to the dimensions of true scientific literacy, Mc Comas and Nouri (2016) measure indicators 5-7 on the dimensions of true scientific literacy according to Shamos (1995). Based on these studies, indicators can be developed that can be used to measure knowledge about the nature of science as one aspect of scientific literacy skills.

Using a review method of more than 20 research results, Rönnebeck et al. (2016) have developed test indicators that measure scientific inquiry skills. Indicators of someone considered competent in scientific inquiry include identifying problems to be investigated, using induction to formulate hypotheses or models, using deduction to produce predictions, and designing experimental procedures. Other parts of the capabilities in scientific inquiry are conducting scientific experiments, observations, or simulations, collecting, organizing, and analyzing data, applying numerical or statistical methods, and explaining unexpected results, as well as using available technology to report, display, and defend results (Lederman et al., 2014; Chu et al., 2021). This inquiry literacy test can measure indicators 1-3 of the dimensions of true scientific literacy, according to Shamos (1995). Based on several studies, indicators can be prepared that can be used to measure scientific inquiry skills as one aspect of scientific literacy skills.

The fourth indicator of true scientific literacy, according to Shamos (1995), namely using logic for induction and deduction, is referred to by Lawson (2018) and Han (2013) as using logic for induction and deduction called scientific reasoning or scientific reasoning thinking. Lawson created a test that measures scientific reasoning skills, called the Lawson Test, while Han (2013) developed a test called the Inventory for Scientific Thinking and Reasoning (iSTAR) Assessment (Bao et al., 2018; Han, 2013). Kind and Osborne (2017) also developed indicators for scientific reasoning based on a review of many studies. According to Kind and Osborne (2017), indicators of scientific reasoning include abilities in mathematical deduction, evaluation of experiments, creating hypothetical models, categorization and classification, probabilistic reasoning, and history-based evolutionary reasoning. Wenning and Vieyra (2020) developed a conceptual framework of scientific reasoning linked to levels of inquiry skills and Bloom's taxonomy. Scientific reasoning, according to Wenning and Vieyra (2020), is a process where logical principles are applied to the scientific process, namely searching for explanations, formulating hypotheses, making predictions, solving problems, creating experiments, controlling variables, analyzing data, developing empirical laws (Wenning & Vieyra, 2020). Based on these studies, indicators can be prepared that can be used to measure scientific reasoning skills as one aspect of scientific literacy skills.

Based on a fairly comprehensive study regarding scientific literacy from Shammos (1995) to Wenning and Vieyra (2020), it can be concluded that there are three dimensions of the meaning of scientific literacy, namely the dimension of scientific reasoning skill, the dimension of scientific inquiry skill and the dimension of the nature of science knowledge. Hanson (2016) explains the three dimensions of scientific literacy as Scientific Reasoning Skill, Scientific Inquiry Skill, and the Nature of Science Knowledge (Hanson, 2016). When combined, the three components can comprehensively explain the profile of students' scientific literacy skills. To find a comprehensive profile of scientific literacy skills, a scientific literacy test instrument is needed to measure all aspects of scientific literacy comprehensively. The novelty of this research is designing a scientific literacy test that is quite comprehensive in terms of content and type of validity, constructing a validity model that uses the Rasch model approach, and administering the test digitally. The types of validity applied in building this test include content, psychometric, construct, and external validity.

There are currently no studies on developing scientific literacy instruments in Indonesia that comprehensively combine the three aspects of scientific literacy dimensions. Rusilowati (2016) developed a scientific literacy test in junior high schools based on science as a body of knowledge, ways of thinking, ways of investigating, and interactions between science, technology, and society (Rusilowati, 2016). Several other researchers have only studied one dimension of scientific literacy but have not connected the three aspects of scientific literacy (Ariyanti et al., 2016; Romine et al., 2017; Wahyuni et al., 2017; Fakhriyah et al., 2017; Rusilowati et al., 2018; Muniroh et al., 2022).

To produce a quality scientific literacy test, technological novelty is needed both in terms of the theoretical approach to the test and in the presentation of the test. Using the Rasch model application as a type of modern test theory is currently considered the most effective approach because it meets all the requirements as an objective measurement model (Bond et al., 2020; Andrich & Marais, 2019). With the help of modern test theory approaches, especially the Rasch model, reporting student skills from these three dimensions can be more effective and meet high assessment validity. In modern test theory measurement, the Rasch model is seen as the most objective measurement model (Sumintono, 2018; Briggs, 2019). The concept of objective measurement in the social sciences and educational assessment, according to Mok and Wright (2004), must have five criteria, namely producing linear measures with equal intervals, an appropriate estimation process, identifying inappropriate items (misfits) or not common (outliers), and can overcome missing data, as well as produce independent measurements of the parameters studied. So far, only the Rasch model can fulfill these five requirements. The Rasch model relates the probability of answering each item correctly (P(i)) as a function of ability (i) with the setting of the difficulty level of item (b).

Presenting tests digitally will have many advantages. These advantages include students' ability to see their exam scores directly, analysis of question items can be done directly, and examiners do not need to make manual corrections or scanning. Likewise, the randomization of questions and answer options makes it difficult for students to work together to complete the test so that cheating behavior can be avoided. Likewise, presenting digital-based tests has become necessary for education in the era of disruption. Therefore, it is necessary to develop a scientific literacy test that can measure students' scientific literacy skills comprehensively using a measurement model based on modern test theory and the use of technology in its application.

This research aims to validate and implement an instrument to measure the scientific literacy of high school science program students, including scientific reasoning skills, scientific inquiry skills, and the nature of science knowledge, or the Lisa3D Test. The valid Lisa3D Test Instrument was then implemented on a broader scale to see the test's external validity and passing limits. These three instruments were built and combined using the Rasch model approach and presented as a digital application to make implementation more effective. It is hoped that the Lisa3D test can become one of the graduation criteria for high school students in the natural science program in Indonesia so that all graduates can master science comprehensively.

METHODS

The research was carried out in a senior high school (SMA) environment where students have medium academic abilities. This research took place from March 2021 to September 2023. This type of research was descriptive survey research with a quantitative approach (Creswell & Creswell, 2017; Gall et al., 2014; Lawrence Neuman, 2019.). Descriptive survey research aims to describe individuals, events, or conditions by studying them as they are. The researcher does not manipulate variables but only describes the sample or variables. This research was a descriptive survey designed to develop and validate an appropriate scientific literacy test to assess the acquisition of scientific literacy for high school students based on scientific reasoning skills, scientific inquiry skills, and the nature of science knowledge called the Lisa3D test. Test development by analyzing test needs and designing tests. The descriptive survey research method was considered appropriate because it would help obtain direct information regarding students› level of scientific literacy skills. The object of this research was an instrument for assessing the scientific literacy of high school science program students, which consists of scientific reasoning skills, scientific inquiry skills, and the nature of science knowledge. It was prepared, revised, and validated using the Rasch model. The product of this research is an instrument that measures scientific literacy skills, which consist of three dimensions: scientific reasoning skills, scientific inquiry skills, and the nature of science knowledge for high school students in the natural science program. This test comprehensively measured the mastery of scientific literacy of high school science program students. It can be used as a graduation criterion to increase the standard of graduation criteria for high school science program students.

At the design stage, research began with analyzing the needs and aspirations of potential users for the tests produced. The next activity was for the researcher to carry out benchmarking related to the three dimensions of scientific literacy tests: compiling and designing test constructs, determining indicators, determining test forms, and developing items. Benchmarking was carried out by reviewing research that had been carried out and consulting with experts who have experience in preparing scientific reasoning tests, scientific inquiry tests, and tests of the nature of scientific knowledge. At the validation stage, researchers validated the test from various aspects, namely content, psychometric, and construct aspects. For the content and psychometric aspects, researchers involved seven resource persons, including material experts and psychometric experts. Content validity and psychometric analyses were carried out qualitatively. The validity of the content aspect was carried out by three professors in the field of science education and one science education practitioner, while the validity of the psychometric aspect was carried out by two professors in the field of education measurement and evaluation and one science education practitioner. The content and psychometric validation models used an inter-rater agreement between the research team and the raters. Based on the agreement of the resource persons and the research team, blueprints for scientific reasoning, scientific inquiry, and the nature of science knowledge tests were produced.

This research used a Rasch modeling approach to analyze the test›s construct validity. The type of construct validity applied is Messick validity (McNamara, 2006), which consists of content aspects, substantive aspects, structural aspects, and external aspects (Ravand & Firoozi, 2016; Bordbar & Alavi, 2021; Newell et al., 2021). The framework for implementing Messick›s construct validity (McNamara, 2006) with the Rasch model approach had been formulated by Susongko (2016), as explained in Table 1. For construct validity, so that item parameter estimates are stable, the instrument was tested in class XII involving 197 students from high school science programs with moderate academic ability. Trials were carried out on 27 July 2022 and 4 August 2022. The test was carried out digitally via the website https://lisatest.id/web/. This system has three roles: administrator, school, and student. Administrators can carry out all test administration and create student and school accounts; create classes of test takers; create tests and test items; plan tests; carry out tests and get access to test results, both total scores and response patterns for each participant. School accounts can only view the test schedule

Table 1. Construct Validity with the Messick Validity Approach and Application of the Rasch Model

Aspects of Construct Validity	Indicator	Criteria
Content	Test item fit (item fit)	P > 0.05
Substantive		P > 0.05
	Person fit statistic	approaching 1.0
Structural	accuracy, sensitivity, and specificity	P< 0,05
External	Invariance Test (LRtest)	approaching 1.0

and access the test results of all students in one school. Student accounts can take tests and view test results.

Instruments validated qualitatively and quantitatively with Rasch modeling were then implemented on target subjects. At the implementation stage, passing standards were determined. A model for reporting scientific literacy scores was carried out, and digital scientific literacy test applications were used. At this stage, a scientific literacy test policy was also formulated to test the external validity of the Lisa3D Test. This stage involved 407 high school students. The tests were conducted on 18 July 2023 and 24 July 2023. The instruments that had been implemented must be tested for external validity. The criteria are learning achievement, national literacy test results, and intelligence tests. If a significant positive correlation is found between the scientific reasoning ability score, the scientific inquiry ability score, and the knowledge score of the essence of science with one of the criteria above, then the scientific literacy test created can be declared to meet the validity of the criteria.

RESULTS AND DISCUSSION

At the stage of preparing the test design, the research began by holding a Focus Group Discussion (FGD) of researchers with all science teachers and several high school principals. This FGD was held on 28 March 2021 and lasted for 4 hours. This FGD aimed to disseminate the theoretical framework for preparing scientific literacy assessment instruments based on scientific reasoning, scientific inquiry, and the nature of science knowledge, as well as to ask for input from teachers and school principals regarding the research plan. Likewise, on 22 April 2021, the researchers asked for input from science education experts regarding the scientific literacy test that would be developed. The results of the FGD can be agreed upon, and the instrument that measures scientific literacy competence consists of three dimensions: scientific reasoning skill, scientific inquiry skill, and the nature of science knowledge for high school science program students. This test will comprehensively measure the mastery of scientific literacy of high school science program students. It can be used as an alternative graduation criterion to increase the graduation standard for high school science program students. This scientific literacy test also reflects the success of science learning in high school, especially in the science program. Through this FGD, it was also agreed to name the literacy test that would be developed as the 3-Dimensional Scientific Literacy Test, abbreviated as the Lisa3D test, to differentiate it from the previously developed scientific literacy test, which still measured basic scientific literacy skills (SLiSIS test) (Susongko et al., 2021). The results of discussions with science education experts provide input so that the scientific literacy tests that are built can approach potential tests so that they can be developed into standard tests.

The scientific literacy test theoretical framework was developed using the scientific literacy test framework according to Shamos (1995) in Hanson (2016). Considering several studies, this theoretical framework was further developed by researchers in three dimensions of scientific literacy: scientific reasoning, scientific inquiry, and the nature of scientific knowledge. The concept of scientific reasoning was formulated based on the theoretical framework developed by Wenning and Vierya (2020) and adapted according to the curriculum context applicable in Indonesia. The difference is that Wenning and Vierya (2015) divide the categories of scientific reasoning into six categories: Rudimentary, Basic, Intermediate, Integrated, Culminating, and advanced, while researchers only limit it to level Culminating. This is due to competency at the level of advanced analysis, which is generally above the ability level of high school students. This researcher's analysis is also supported by the results of Hanson's (2016) research, which measured scientific reasoning skills at the high school level only at the level of Culminating. A total of 25 scientific reasoning test indicators have been produced in the research.

The theoretical framework for developing instruments for the dimensions of scientific inquiry is based on the framework developed by Wenning (2007) (Wenning & Vierya, 2020). Wenning (2007) builds scientific inquiry skills such as Discovery Learning, Interactive Demonstrations, Inquiry Lessons, Inquiry Labs, and Hypothetical Inquiry. Researchers compiled six scientific inquiry test skills and indicators using scientific narratives appropriate to the context of Indonesian students. The theoretical framework for measuring tests of the nature of science knowledge is derived from competencies formulated as indicators of someone understanding the nature of science according to several previous research as explained by Suzuri-Hernandez (2010). Researchers added one indicator based on this theoretical framework: students' ability to differentiate between fact and belief phenomena. This is very important because the ability to distinguish facts and beliefs is the main indicator of the difference between science and non-science. Thus, there are a total of eight indicators. A complete study of the conceptual framework of the three-dimensional scientific literacy test has been published by researchers. This study discusses the competency of the three-dimensional scientific literacy test up to the reduction of the indicators (Susongko et al., 2022). The results of the construct validity analysis for the Scientific Reasoning Test, Scientific Inquiry Test, and Nature of Science Test can be seen in Table 2. The difficulty level of items from the three valid tests can be seen in Table 3-Table 5.

Table 2. Results of Construct Validity Analysis Using the Messick Validity Approach and the Rasch Model Application to Scientific Reasoning Test Items

Aspects of Construct Validity	Scientific Reasoning	Scientific Inquiry	Nature of Science
Content	A total of 44 (88%) test items matched the model	A total of 12 (100%) test items matched the model	A total of 15 (93.7%) test items matched the model
Substantive	A total of 176 (90.7%) student responses matched the model Accuracy: 0.744 Sensitivity: 0.798 Specificity: 0.68	A total of 182 (93.8%) student responses matched the model Accuracy: 0.797 Sensitivity: 0.894 Specificity: 0.597	A total of 189 (97.4%) student responses matched the model Accuracy: 0.732 Sensitivity: 0.829 Specificity: 0.59
Structural	The inconsistent item is item no. 2	Inconsistent items are numbers 10,11 and 12	The inconsistent item is item number 7
External	0.787	0.5088	0.4663

Table 3. Level of Difficulty of Scientific Reasoning Test Details

No	Level Difficulty	No	Level Difficulty	No	Level difficulty	No	Level difficulty	No	Level difficulty
1	- 1.402	6	2.345	11	0.255	16	0.121	21	0.591
2	-0.084	7	-0.130	12	1.244	17	0.233	22	0.941
3	-0.569	8	1.325	13	-0.729	18	-0.038	23	-0.343
4	-1.564	9	1.089	14	2.162	19	-1.185	24	0.917
5	-0.595	10	1.649	15	0.008	20	1.114	25	1.271

The Scientific Reasoning test items in Table 3 can be seen in Arfiani et al. (2023). The Scientific Inquiry items as in Table 4 can be seen at https://bit.ly/Inquiry-2024-UPS.

Table 4. Level of Difficulty of Scientific Research Test Details

No	Difficulty Level	No	Difficulty Level
1	-0.170	4	2.061
2	-0.806	5	-0.170
3	0.512	6	-0.980

Table 5. Level of Difficulty of Nature of Science Test Details

No	Difficulty Level	No	Difficulty Level
1	0.099	5	0.937
2	0.648	6	0.277
3	0.517	7	0.914
4	-0.614	8	0.364

See Susongko et al. (2024) to learn more about the nature of science test items.

As a standard proficiency test, the Lisa3D Test must have minimum criteria where a student is declared to have minimal proficiency. Determining minimum criteria is what is called standard settings. Determining passing standards is a procedure for determining the minimum limit of competency or ability that a test taker must have after participating in a learning program. In this case, the learning program is as long as students participate in high school science programs. In determining the passing limit for the Lisa3D test, the Bookmark Method with the Rasch model approach is used. The Bookmark method is widely used because it is more practical, can be used in various tests, and is easy to implement. Likewise, the Bookmark method provides more accurate results because it can apply the use of modern test theory (IRT or Rasch model), uses ordered item data, and is more efficient in determining the passing score on the test based on the results of expert discussions (Lewis et al., 2012; Clauser et al., 2017; Baldwin et al., 2020). The Bookmark method is used to conduct analysis using ordered test items. Test items are ordered from the easiest test items to the most challenging test items. The assessors divide ability levels, namely basic, intermediate, and advanced levels, or with other categorizations. There is a comparison of the assessment of the level of item difficulty between the results of the analysis and the assessments of the judges (raters). Paying attention to the average difficulty level of the 50 scientific reasoning test items means a passing limit for the scientific reasoning aspect can be determined at 56% of the ideal achievement. In the same way, a passing limit for scientific inquiry aspects can be produced at 50% of the ideal achievement, while for the nature of science aspect, it is determined at 56.25% of the ideal score.

Based on the results of implementation involving two schools, each with 195 School A students and 212 School B students, School B students performed better than School A students for all aspects of scientific literacy. In the scientific reasoning aspect, there is a striking difference where the number of School B students who pass is almost 17 times compared to the number of School A students who pass. In scientific inquiry, the achievements of both schools are balanced, exceeding 70% of students who have exceeded the passing limit. Regarding the nature of science knowledge, the number of students who pass from School B is 10% higher than those who pass from School A. Detailed passing percentages can be seen in Table 6.

Table 6. Passing Percentage of Students Who Take the Three-Dimensional Scientific Literacy Test by School

Sub Test	Passing percentage (%) (School A)	Passing percentage (%) (School B)
Scientific Reasoning	2.56	23.58
Scientific Inquiry	72.36	74.36
Nature of Science	42.97	52.10

Based on the analysis of each item, the scientific reasoning achievements of School A students are generally lower than those of School B students except for three indicators, namely the ability to generalize, the ability to use conditional thinking, and the ability to use proportional thinking to make a decision. The achievement of scientific reasoning in both schools is highest at the elementary level, namely at School A at 50.77%, while at School B students, it is 48.58%. However, School B students are superior in terms of level attainment. 3.77% of students are at the integrated level, and 25% are at the intermediate level. As a comparison, measurements of reasoning carried out by Hanson (2016) in America on high school students gave an average result of 30%. By using a three-dimensional scientific literacy test in the aspect of scientific reasoning, it can be seen that the achievement of students at School A is 32.76%, while for School B students, it is 41.9%. Theoretically, we expect high school students' reasoning skills to reach a high level, but empirically, the average achievement is at a basic level. However, among School B students, 25% are at the intermediate level, and 3.77% are at the integrated level.

Table 7. Three-dimensional Scientific Literacy Skill in the Scientific Reasoning Aspect

Scientific Reasoning Skill
Classify, Conceptualize, Conclude, Contextualize
Generalize, Sequence, Formulate problems, Estimate, Explain
Predict, Use conditional thinking, Apply Information
Explain Relationships, Use Quantitative Data, Use Combinatorial Thinking, Use Relationship Thinking, Define Problems in a Study, Define problems that can be solved, Design and conduct controlled scientific investigations
Interpret quantifiable data to establish laws
Determine whether the answer to the problem or question is rational
Summarize to justify conclusions based on empirical evidence
Use causality thinking to differentiate from coincidence cause and effect
Use correlational thinking to differentiate from coincidence cause and effect
Use proportional thinking to make a decision

(Hanson, 2016)

Likewise, in scientific inquiry, almost all indicators of School B students' achievements are higher than those of School A students, except for identifying scientific principles and their relationships. In this indicator, the achievement of School B students is 47.59%, while the achievement of School A students is 58.94%. Except for this indicator, student achievement in both schools in the other five indicators is above 50% as the passing limit. This shows that scientific inquiry skills on indicators to identify scientific principles and their relationships need to be improved.

Regarding knowledge of the nature of science, almost all indicators of School B students' achievements are higher than those of School A students, except for distinguishing between facts and explanations. In this indicator, the achievement of School B students was 49.04%, while the achievement of School A students was 56.21%. There are only two indicators where student achievement at both schools has exceeded the graduation limit, namely 56.25%. These indicators are: (1) Understanding that observations are influenced and guided not only by scientific theory but also by the beliefs, values, attitudes, commitments, training, previous knowledge, past experiences, and expectations of a scientist and (2) Understanding that Providing an explanation is not just a matter of collecting data and doing it logically to get it, but intuition, imagination, and creativity are needed. Table 7 shows the profile of students' three-dimensional scientific literacy skills for aspects of scientific reasoning.

Meanwhile, for scientific literacy skills, the scientific inquiry aspect includes building basic concepts from experience involved in explanations and making predictions that enable teachers to obtain, identify, confront, and resolve alternative conceptions. Another skill is identifying scientific principles or relationships and finding empirical data based on variable measurements. The scientific inquiry aspect also measures a person's ability to solve problems related to authentic situations when working individually or in groups using problem-based and project-based approaches with collaborative principles and producing more realistic explanations for observed phenomena. Meanwhile, for scientific literacy skills, the aspect of the nature of science knowledge includes the ability to differentiate between facts and explanations and understand that observations are influenced and guided not only by scientific theory but also by beliefs, values, attitudes, commitment, training, previous knowledge, past experiences, and the hopes of a scientist. The aspect of the nature of science also includes an understanding that providing an explanation is not just a matter of collecting data but requires intuition, imagination, and creativity, as well as an understanding that the scientific method is a strategy for concluding explanations of phenomena, including evaluating comparisons of specific predictions from observational data with existing data from the real world. Another aspect of the nature of science knowledge is understanding that explanations are not determined only by data but by several different explanations and understanding that scientific explanations must meet several criteria, one of which is observational confirmation of predictions that have been made. Ultimately, the aspect of the nature of science knowledge also includes the ability to understand that a scientific explanation can be abandoned or even rejected in favor of a new, better explanation and to be able to distinguish between facts and beliefs.

The instruments used in the target population must be evaluated on various things, such as the validity of the scores, the quality of test implementation, and the usefulness of the test program. From these various aspects, score validity must be carried out. This is because the validity of the score is related to external validity, which is one aspect of validity that must be fulfilled in preparing the test. The Lisa3D test score results at this implementation stage will be compared with other more valid scores to measure its external validity. The analysis results show that the correlation between scientific reasoning scores and mathematics learning achievement scores is 0.296, and the correlation between scientific reasoning scores and physics learning achievement scores is 0.697. By testing the significance of 0.01 with the help of SPSS software, the analysis results show that both correlations are significant, so it can be stated that there is a real and significant correlation between scientific reasoning scores and mathematics learning achievement scores. Likewise, it can be stated that there is a real and significant correlation between scientific reasoning scores and physics learning achievement scores.

Until now, the development of a scientific literacy test in Indonesia has never been achieved by combining the three competencies as was done in this research, namely scientific reasoning, scientific inquiry, and knowledge of the nature of science. As a reference for the development of scientific literacy assessments in Indonesia, PISA standards are still widely used and are related to scientific literacy skills at the junior high school level (Rosana et al., 2020; Atta & Ara, 2019;

Ni'mah, 2019; Ardiyanti et al., 2019). Some scientific literacy researchers in Indonesia have built scientific literacy tests based on the subject matter (Adnan et al., 2021; Mulbar & Bahri, 2021; Rusilowati et al., 2018; Novaristiana et al., 2019; Dewi et al., 2019). Thus, the research conducted has produced novelty related to several new things produce. First, the built scientific literacy combines three basic skills (scientific reasoning, scientific inquiry, and knowledge of the nature of science) as a standard for comprehensive scientific literacy mastery. Second, the developed scientific literacy was applied to high school students, where scientific literacy assessments in high school have rarely been carried out. Third, scientific literacy test is built based on modern test theory, the Rasch Model, and test validity using Messick Validity (2006), a reasonably comprehensive validity theoretical framework. In modern test theory measurement, the Rasch model is seen as the most objective measurement model (Sumintono, 2018; Briggs, 2019).

From a psychometric perspective, potential tests will be more valid than actual tests. This is because potential tests can measure a person's abilities before experiencing learning or are latent, which can be caused by learning over a long period. These potential tests, for example, are general intelligence tests and aptitude tests. Actual tests are produced as a result of a learning program. The Lisa3D test is basically an actual test even though the abilities measured are built over a long time, from learning experiences in middle and high school, perhaps even knowledge outside the classroom. In this case, concurrent or same-time validity, information on potential tests that students have carried out is required to measure external validity. Unfortunately, researchers do not have information regarding general intelligence (IQ) tests or aptitude tests students have carried out. For this reason, researchers took students' learning achievement scores in the even semester of the 2022/2023 academic year when students were in class XI. The scores used are Mathematics and Physics learning achievement scores. This external validity will be fulfilled if the correlation between the scores in the Lisa3D test and the learning achievement scores is real and significant (Veiga, 2016; Kluge et al., 2017).

Of the three aspects of scientific literacy, only scientific reasoning can be tested for criterion validity because the reference scores are limited, in this case, only mathematics and physics learning achievement. Several research studies show that scientific reasoning greatly influences student performance in mathematics and physics subjects (O'Connell, 2018; Ramos et al., 2013; Mayer et al., 2014). Thus, student performance in mathematics and physics subjects, shown by mathematics and physics learning achievement scores, can be used as a criterion score to see the validity of scientific reasoning test scores. The results of this research show a real and significant correlation between scientific reasoning scores and mathematics learning achievement scores, and there is a real and significant correlation between scientific reasoning scores and physics learning achievement scores. Thus, the external validity of the scientific reasoning aspect is fulfilled.

In aspects of scientific inquiry and the nature of science in the Lisa3D test, until now, researchers have not obtained a reference score previously, followed by students who took the Lisa3D test. To test external validity, the researchers have tested the correlation between the achievements of each indicator in the two tests in two schools. Assuming that if the two indicators of the scientific inquiry test or the nature of science knowledge test are significantly correlated, then it can be concluded that the test meets external validity. This can be done because score consistency is a measure of evidence validity. Evidence validity can be seen from the consistency of two score distributions that measure the same content (Sireci & Faulkner-Bond, 2014; Sireci & Benítez Baena, 2023). The analysis results show that the correlation between the scientific inquiry scores for the two high school students is 0.926, while for knowledge of the nature of science, it is 0.635. From this information, it can be stated that the Lisa3D test in scientific inquiry and the nature of science meets external validity.

The development of high-stakes tests in the atmosphere of implementing emancipated learning in Indonesian schools is contrary to Government policy. Researchers can feel this because the students who took the test were not completely serious about taking the test. This is natural because the attempted test did not provide any consequences for students, so motivation to succeed on the Lisa3D test was low. Low motivation will result in less than optimal learning achievement test results (Filgona et al., 2020; Howard et al., 2021). Moreover, the emancipated curriculum has no standard test for students to graduate after taking a certain level of school. However, with socialization and motivation, students can understand the importance of mastering scientific literacy skills to build scientific skills in subsequent education.

CONCLUSION

This research has produced a standard instrument for measuring the scientific literacy of high school students in science programs based on three dimensions: scientific reasoning, scientific inquiry, and knowledge of the nature of science, or called the 3-Dimensional Scientific Literacy Test/Lisa3D Test. The standard test consists of 25 scientific reasoning test items, six scientific inquiry test items, and eight items of knowledge of the nature of science. All these tests have met all validity, including content, psychometric, construct, and external validity. The validity of the construct applied consists of content, substantive, structural, and external aspects. The standard test has been successfully implemented through the lisatest.id page with a cut of 56% for the scientific reasoning test, 50% for the scientific inquiry test, and 56.25% for the test of knowledge of the nature of science. The Lisa3D test can be used as a standard instrument to measure scientific literacy skills in senior secondary schools in natural science programs. Lisa3D test results can be used as additional passing criteria for high school students. In this way, high school graduates can be guaranteed comprehensive mastery of scientific literacy.

ACKNOWLEDGMENT

The researchers would like to thank the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for funding this research for three years from 2021 to 2023 with contract number 1/061013/PG/SP2H/JT/2021;139.a/A-5/LPPM-UPS/VII/2022; 303.a/K/A-5/LPPM/UPS/IV/2023. Likewise, the researchers would like to thank all the teachers, students, and high school principals who have been actively involved and helped carry out this research and the Research and Service Institute Universitas Pancasakti Tegal for assisting in administrating this research process.

REFERENCES

Adnan, A., Usman, U., & Bahri, A. (2021). Scientific literacy skills of students: Problem of biology teaching in junior high school in south sulawesi, indonesia. *International Journal of Instruction*, *14*(3), 847-860.

Andrich, D., & Marais, I. (2019). A course in Rasch measurement theory. Measuring in the educational, social and health sciences, 41.

Anselmi, P., Vidotto, G., Bettinardi, O., & Bertolotti, G. (2015). Measurement of change in health status with Rasch models. Health and Quality of Life Outcomes, 13, 1-7.

Ardiyanti, Y., Suyanto, S., & Suryadarma, I. G. P. (2019, October). The role of students science literacy in Indonesia. In Journal of Physics: Conference Series (Vol. 1321, No. 3, p. 032085). IOP Publishing.

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.

Ariyanti, A. I. P., Ramli, M., & Prayitno, B. A. (2016). Preliminary Study on Developing Science Literacy Test for High School Students in Indonesia. In *International Conference on Teacher Training and Education*. Sebelas Maret University.

Atta, H. B., & Aras, I. (2020). Developing an instrument for students scientific literacy. In *Journal of Physics: Conference Series* (Vol. 1422, No. 1, p. 012019). IOP Publishing.

Baldwin, P., Margolis, M. J., Clauser, B. E., Mee, J., & Winward, M. (2020). The choice of response probability in bookmark standard setting: an experimental study. Educational Measurement: Issues and Practice, 39(1), 37-44.

Bao, L., Xiao, Y., Koenig, K., & Han, J. (2018). Validity evaluation of the Lawson classroom test of scientific reasoning. *Physical Review Physics Education Research*, *14*(2), 020106.

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

Bordbar, S., & Alavi, S. M. (2021). Detecting gender-biased items in a high-stakes language proficiency test: using Rasch model measurement. International Journal of Quantitative Research in Education, 5(3), 277-310.

Briggs, D. C. (2019). Interpreting and visualizing the unit of measurement in the Rasch Model. Measurement, 146, 961-971.

Chu, S. K. W., Reynolds, R. B., Tavares, N. J., Notari, M., & Lee, C. W. Y. (2021). 21st century skills development through inquiry-based learning from theory to practice. Springer International Publishing.

Clauser, B. E., Baldwin, P., Margolis, M. J., Mee, J., & Winward, M. (2017). An experimental study of the internal consistency of judgments made in bookmark standard setting. Journal of Educational Measurement, 54(4), 481-497.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Dewi, C. A., Khery, Y., & Erna, M. (2019). An ethnoscience study in chemistry learning to develop scientific literacy. *Jurnal Pendidikan IPA Indonesia*, *8*(2), 279-287.

Erduran, S., Dagher, Z. R., Erduran, S., & Dagher, Z. R. (2014). Reconceptualizing nature of science for science education (pp. 1-18). Springer Netherlands.

Fakhriyah, F., Masfuah, S., Roysa, M., Rusilowati, A., & Rahayu, E. S. (2017). Student's science literacy in the aspect of content science?. Jurnal Pendidikan IPA Indonesia, 6(1).

Filgona, J., Sakiyo, J., Gwany, D. M., & Okoronka, A. U. (2020). Motivation in learning. Asian Journal of Education and social studies, 10(4), 16-37.

Fives, H., Huebner, W., Birnbaum, A. S., & Nicolich, M. (2014). Developing a measure of scientific literacy for middle school students. Science Education, 98(4), 549-580.

Gall, M. D., Gall, J. P., & Borg, W. R. (2014). Applying educational research: How to read, do, and use research to solve problems of practice. Pearson Higher Ed.

Han, J. (2013). Scientific reasoning: Research, development, and assessment (Doctoral dissertation, The Ohio State University).

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

Howard, J. L., Bureau, J. S., Guay, F., Chong, J. X., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. Perspectives on Psychological Science, 16(6), 1300-1323.

Kind, P. E. R., & Osborne, J. (2017). Styles of scientific reasoning: a cultural rationale for science education?. Science Education, 101(1), 8-31.

Kluge, F., Gaßner, H., Hannink, J., Pasluosta, C., Klucken, J., & Eskofier, B. M. (2017). Towards mobile gait analysis: concurrent validity and test-retest reliability of an inertial measurement system for the assessment of spatio-temporal gait parameters. Sensors, 17(7), 1522.

Lawrence Neuman, W. (2019). Metodologi Penelitian sosial; Pendekatan Kualitatif dan Kuantitatif. Jakarta : PT Indeks

Lederman, J. S., Lederman, N. G., Bartos, S. A., Bartels, S. L., Meyer, A. A., & Schwartz, R. S. (2014). Meaningful assessment of learners' understandings about scientific inquiry—The views about scientific inquiry (VASI) questionnaire. Journal of research in science teaching, 51(1), 65-83.

Lewis, D. M., Mitzel, H. C., Mercado, R. L., & Schulz, E. M. (2012). The bookmark standard setting procedure. In Setting performance standards (pp. 225-253). Routledge

Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. Learning and Instruction, 29, 43-55.

McComas, W. F. (2013). Nature of science in the science curriculum and in teacher education programs in the United States. In International handbook of research in history, philosophy and science teaching (pp. 1993-2023). Dordrecht: Springer Netherlands.

McComas, W. F., & Nouri, N. (2016). The nature of science and the next generation science standards: Analysis and critique. Journal of Science Teacher Education, 27(5), 555-576.

McNamara, T. (2006). Validity in language testing: The challenge of Sam Messick's legacy. *Language Assessment Quarterly: An International Journal*, *3*(1), 31-51.

Molenda, M. (2015). In search of the elusive ADDIE model. *Performance Improvement*, *54*(2), 40-42

Mulbar, U., & Bahri, A. (2021, February). Biology Science Literacy of Junior High School Students in South Sulawesi, Indonesia. In *Journal of Physics: Conference Series* (Vol. 1752, No. 1, p. 012084). IOP Publishing.

Muniroh, N., Rusilowati, A., & Isnaeni, W. (2022). Instrument Development of Science Literacy Assessment with Socio-Sciences Contains in Natural Science Learning for Elementary School. *Journal of Research and Educational Research Evaluation*, *11*(1), 15-22.

Newell, S. J., Chur-Hansen, A., & Strelan, P. (2021). A construct validity analysis of the concept of psychological literacy. Australian Journal of Psychology, 73(4), 1922069.

Ni'mah, F. (2019). Research trends of scientific literacy in Indonesia: Where are we?. *Jurnal Inovasi Pendidikan IPA*, *5*(1), 23-30.

Novaristiana, R., Rinanto, Y., & Ramli, M. (2019). Scientific literacy profile in biological science of high school students. JPBI (jurnal pendidikan biologi indonesia), 5(1), 9-16.

O'Connell, M. (2018). The power of cognitive ability in explaining educational test performance, relative to other ostensible contenders. Intelligence, 66, 122-127.

Osborne, J. (2023). Science, scientific literacy, and science education. In Handbook of research on science education (pp. 785-816). Routledge.

Ramos, J. L. S., Dolipas, B. B., & Villamor, B. B. (2013). Higher order thinking skills and academic performance in physics of college students: A regression analysis. International Journal of Innovative Interdisciplinary Research, 4(1), 48-60.

Ravand, H., & Firoozi, T. (2016). Examining construct validity of the master's UEE using the Rasch model and the six aspects of the Messick's framework. *International Journal of Language Testing*, *6*(1), 1-18.

Romine, W. L., Sadler, T. D., & Kinslow, A. T. (2017). Assessment of scientific literacy: Development and validation of the Quantitative Assessment of Socio-Scientific Reasoning (QuASSR). *Journal of Research in Science Teaching*, *54*(2), 274-295.

Rönnebeck, S., Bernholt, S., & Ropohl, M. (2016). Searching for a common ground–A literature review of empirical research on scientific inquiry activities. Studies in Science Education, 52(2), 161-197.

Rosana, D., Widodo, E., Setianingsih, W., & Setyawarno, D. (2020). Developing assessment instruments of PISA model to measure students' problem-solving skills and scientific literacy in junior high schools. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, *8*(2), 292-305.

Rusilowati, A., Kurniawati, L., Nugroho, S. E., & Widiyatmoko, A. (2016). Developing an Instrument of Scientific Literacy Assessment on the Cycle Theme. *International Journal of Environmental and Science Education*, *11*(12), 5718-5727.

Rusilowati, A., Nugroho, S. E., Susilowati, E. S. M., Mustika, T., Harfiyani, N., & Prabowo, H. T. (2018, March). The development of scientific literacy assessment to measure student's scientific literacy skills in energy theme. In *Journal of Physics: Conference Series* (Vol. 983, No. 1, p. 012046).

Rusilowati, A., Nugroho, S. E., Susilowati, E. S. M., Mustika, T., Harfiyani, N., & Prabowo, H. T. (2018, March). The development of scientific literacy assessment to measure student's scientific literacy skills in energy theme. In *Journal of Physics: Conference Series* (Vol. 983, No. 1, p. 012046). IOP Publishing.

Sireci, S. G., & Benítez Baena, I. (2023). Evidence for test validation: A guide for practitioners. Psicothema.

Sireci, S. G., & Faulkner-Bond, M. (2014). Validity evidence based on test content. Psicothema.

Sumintono, B. (2018, February). Rasch Model Measurements as Tools in Assesment for Learning. In 1st International Conference on Education Innovation (ICEI 2017). Atlantis Press

Susongko, P , Kusuma, M, Yuenyong, C., & Arfiani, Y. (2024). Developing the Nature of Science (NOS) Test Using Covid-19 Theme. *Pegem Journal of Education and Instruction*, *14*(2), 57-66.

Susongko, P. (2016). Validation of science achievement test with the rasch model. Jurnal Pendidikan IPA Indonesia, 5(2), 268-277.

Susongko, P. U. R. W. O., Arfiani, Y., & Kusuma, M. (2021). Determination of Gender Differential Item 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., & Arfiani, Y. (2022, December). 3-dimensional scientific literacy assessment framework for senior high school science program students. In AIP Conference Proceedings (Vol. 2600, No. 1). AIP Publishing.

Susongko, P., Kusuma, M., Arfiani, Y., Samsudin, A., & Aminudin, A. (2020). Revising of the Integrating Scientific Literacy Skills Scale (ISLS) with Rasch Model Analysis. *Journal for the Education of Gifted Young Scientists*, *8*(4), 1583-1602.

Susongko, P., Widiatmo, H., Kusuma, M., & Afiani, Y. (2019). Development of Integrated Science-Based Science Literacy Skills Instruments using the Rasch Model. *Unnes Science Education Journal*, *8*(3).277-292

Suzuri-Hernandez, L. J. (2010). *Exploring school students' views of the nature of science* (Doctoral dissertation, University of York).

Veiga, F. H. (2016). Assessing student Engagement in School: Development and validation of a four-dimensional scale. Procedia-Social and Behavioral Sciences, 217, 813-819.

Wahyuni, S., Indrawati, I., Sudarti, S., & Suana, W. (2017). Developing science process skills and problem solving abilities based on outdoor learning in junior high school. *Jurnal Pendidikan IPA Indonesia*, *6*(1).

Wenning, C. J., & Vieyra, R. E. (2020). Teaching High School Physics: The Nature of Physics Teaching. USA: AIP Publishing