Developing Cognitive Instruments Based on Science Literacy on Sound Materials

Mustika Wati, Asmaul Husna, Sarah Miriam, Saiyidah Mahtari

Physics Education Study Program
Universitas Lambung Mangkurat
Banjarmasin, Indonesia
asmaulhusna.comz@gmail.com

Abstract— **This study aimed to produce cognitive instruments that could help teachers train students' literacy skills. The development of cognitive instruments based on scientific literacy on sound material was aimed at; (1) Describe the validity of the instrument being developed, (2) Describe the reliability of the instrument being developed, (3) Describe the level of difficulty of the instrument being developed, and (4) Describe the distinguishing power of the instrument being developed. This research was a research and development with research design using the development of Research and Development (R & D) adaptation of Borg and Gall (1983), namely potential and problems, data collection, product design, design validation, product testing, product revision, usage testing, and final product revisions. The resulting product was a cognitive instrument based on scientific literacy in the form of a description test with a level of C2 to C4 of 10 questions. The products were tested in Banjarmasin State Junior High School 4, Banjarmasin State Middle School 15, and Banjarmasin State Middle School 24. The data analysis technique uses item response theory, the Rasch model. The results showed: (1) the validity of the instrument developed was valid, (2) the reliability of the instrument developed was excellent with alpha Cronbach 0.86, (3) the difficulty level of the instrument developed was very difficult 20%, difficult 30%, easy 30% and very easy 20%, and (4) the differentiating power of the instruments developed was 80% good and 20% bad. The conclusion of cognitive literacy-based science instrument with good category was obtained so that it is feasible to use.**

Keywords— Cognitive Instruments, Science Literacy, Sound

I. INTRODUCTION

Education is an important aspect in the development of human resources and is a determinant of the success of development in improving and developing the quality of human resources [1]. By improving the quality of learning and the quality of the assessment system can improve the quality of education

In the teaching and learning process, the assessment factor of learning outcomes is often not noticed, in the sense that the teacher only looks at and looks at the learning presentation only. Learning takes place well, the practicum runs well, but at the time of making the assessment instrument, it does not pay attention to the learning objectives to be achieved. As a result, the learning outcomes test is made inappropriate.

Assessment instruments are said to be good or meet the requirements can be accounted for in terms of validity, reliability, level of difficulty, and distinguishing power [2]. Based on the results of interviews with science subject teachers conducted at Banjarmasin State Middle School 24, Banjarmasin State Middle 4, and Banjarmasin State Middle School 15, information was obtained that for cognitive assessment, the teacher used instruments in the form of questions made by the Banjarmasin City Education Office as a result of Teacher Subjects (MGMP) IPA, while for daily tests made by the teacher itself. The instrument made by the teacher has never been applied to the item analysis before, meaning that the instrument received by students does not necessarily meet the criteria for making good and correct assessment instruments.

Based on interviews from the three schools observed, it was found that there were already schools that applied literacy but had not been integrated into the problem. In addition, the daily test questions are often multiple choice questions. In multiple choice questions, the teacher only assesses the right and wrong answers; the teacher misses the opportunity to assess students' ability to read science and communicate science freely and creatively.

Science literacy is a person's competence in deepening and communicating science and applying scientific knowledge in solving problems so that they have high attitudes and vulnerabilities in themselves and their environment when making decisions in accordance with scientific considerations [3]. Understanding science learning that leads to the formation of scientific literacy of students in Indonesia still seems not fully understood by science teaching teachers. From the results of the 2015 PISA tests and evaluations, it shows that the scientific literacy abilities of Indonesian students are still relatively low. In a row, the average score of achievement of Indonesian students for science, reading, and mathematics was ranked 62, 61, and 63 of the 69 countries evaluated. Students are generally poorly trained in solving problems with characteristics such as questions that apply scientific literacy [4]. In addition, the low scientific literacy abilities of students are also caused by the familiarity of students in science learning who ignore the importance of the ability to read and write science as a must-have competency.

One of the things that causes the low level of scientific literacy of students is the small desire and motivation of

learning of students towards science, the embeddedness of scientific literacy in students, it will make students become people who are able to explore the material and characteristics of science and are able to apply scientific concepts in daily life [5]. The teacher's knowledge of scientific literacy is still very much limited so that it causes application when learning takes place is still not maximum [6].

The two problems above, namely the unclear quality of cognitive instruments used in schools and the low scientific literacy of students and teachers in Indonesia, need to be dealt with seriously. One of the ways to overcome the above problems is by making an instrument that is in accordance with the learning objectives and has been tested for validity, reliability, level of difficulty, and the power of dissemination, as well as familiarizing students to practice scientific literacy.

Development of instruments based on scientific literacy needs to be developed to be able to assess the process and results of student learning and encourage students to understand the nature of science better comprehensively and can stimulate all students to be active in all learning processes [6]. This form of the test requires the ability of students to convey, choose, compile, and integrate ideas or ideas that they already have by using their own words. By giving a question in the form of reading, it will help students to add ideas and ideas that are on their mind [2]

Several studies have been conducted with regard to the quality of cognitive tests and scientific literacy, including references [1] suggesting that instruments need to be developed which have known construct validity and quality to overcome the problem of cognitive test devices that have not known construct validity and quality. Science literacy can be used to see students' ability to apply scientific knowledge to solve problems.

Based on the description above, efforts are needed to develop cognitive instruments to assess student learning outcomes. The purpose of this study was to produce a quality of cognitive instruments based on scientific literacy on sound material that is of high quality and feasible to use.

II. METHOD

This research was a research and development with research design using research and development (R & D) development procedures with procedural models from Borg and Gall adopted [7]. This research was conducted in class VIIA, and VIIIC in Banjarmasin State Middle School 24, VIID and VIIIE classes at Banjarmasin State Junior High School 4, and VIIIE, VIIIF, and VIIIH classes in Banjarmasin State Middle School 15 were conducted from March 2017 - January 2018.

Data collection techniques were in the form of tests and non-tests. The non-test technique was by interviewing, while the test technique was by using description questions at the time of evaluation of learning outcomes. The technique of analyzing data was Rasch modeling with item response theory. Analysis of the validity of the question was based on expert test results, product trial results, and Rasch programs.

Validated product design produced qualitative data in the form of assessments, responses, and suggestions on each item. The calculation results were then interpreted using expert validity criteria [8]. Consistency or compatibility of the validation results from the three validators, then alpha Cronbach reliability was used, then adjusted to the criteria for the assessment aspects determined based on Table 1.

TABLE I.   INSTRUMENT RELIABILITY CRITERIA

Reliability Coefficient	Criteria
$0,80 \leq r$	High reliability
$0,40 \leq r \leq 0,80$	Moderate reliability
$r < 0,40$	Low reliability

Validation analysis of product trial results was calculated using the product moment formula with rough numbers. The calculation results were then interpreted using the correlation coefficient criteria [9].

Analysis of the validity of the question was done using the Rasch program. In this program, a question is valid depending on the MNSQ value, the ZSTD value, and the measurement correlation value produced.

The criteria for MNSQ values could be interpreted using MNSQ value criteria. If the MNSQ value was more than 2.0, then the implication was reducing the quality of the measurement system, if the MNSQ value was in the range of 1.5 to 2.0, then the implications were not good for instrument making but do not reduce quality, if the MNSQ value was in the range of 0.5 to 1.5, there were good quality implications for measurement, whereas if the MNSQ value was less than 0.5, the implications were less productive for measurement but did not reduce quality and allow high-reliability errors [10]. The value of ZSTD can be interpreted using Table 2.

TABLE II.   INTERPRETATION OF ZSTD VALUE

ZSTD Value	Implications for Measurement
$\geq 3,0$	Data is not expected if it matches the model. But with a large sample, the mismatch may be smaller.
$2,0 – 2,9$	Data seems unpredictable.
$-1,9 – 2,9$	Data has logical estimates.
≤ -2	Data is too predictable.

Criteria for measuring the correlation are located between the ranges of values from 0.4 to 0.85. Values can be said to be valid and appropriate if there is no MNSQ, ZSTD, and correlation measurement values that are outside the criteria simultaneously [10].

The reliability of the questions can be analyzed using the Cronbach Alpha value. The criteria for interpreting Cronbach's Alpha values can be interpreted using Table 3.

TABLE III. INTERPRETATION OF ALPHA CRONBACH VALUE

Criteria	Alpha Cronbach
Bad	< 0,5
Poor	0,5 – 0,6
Moderate	0,6 – 0,7
Good	0,7 – 0,8
Excellent	> 0,8

Analysis of problem difficulty level based on a design test result can be calculated by using a formula based on [11]. If the difficulty value obtained is greater than or equal to 0.00 and less than or equal to 0.30, it is in the difficult category, if the difficulty value is greater than 0.30 and less than or equal to 0.70, then is in the middle category, whereas if the difficulty value is greater than 0.70 and less than or equal to 1.00, it is in an easy category [11].

The difficulty level of the problem is analyzed using the number of logit contained in the questionnaire column when the number of logit is bigger, then the level of problem difficulty will also increase. The criteria that can be used to interpret difficulty levels can be seen in Table 4.

TABLE IV. RASCH DIFFICULTY LEVEL CRITERIA

Kriteria	Interpretation
$> 0,36$	Very difficult
$0,36 \geq TK > 0,00$	Difficult
$0,00 \geq TK \geq -0,36$	Easy
$< -0,36$	Very Easy

The distinguishing index can be obtained by calculation using a formula based on [11]. If the value of differentiator is more than or equal to 0.40 and less than or equal to 1.00, then the question can be accepted, if the value of the differentiator is more than or equal to 0.30 and less than 0.40, then the question can be accepted, but it needs to be corrected, if the value of the differentiator is more than or equal to 0.20 and less than 0.30, then the question needs to be fixed, whereas if the value of the differentiator is more than or equal to 0.00 and less than 0.20, then the question not used [11].

III. RESULT AND DISCUSSION

Cognitive literacy-based instruments developed have been validated and classed to produce instruments that are feasible to use. The results of the cognitive instrument validation by three validators can be seen in Table 5.

TABLE V. EXPERT VALIDATION

Validator	Mean
Validator 1	2,9
Validator 2	3,3
Validator 3	4,0
Mean	3,4
Reliability	0,98

Table 5 shows that the results of instrument validation with an average of 3.4 in the category are very good and the reliability value is 0.98 in the high category. For general construction, aspects consist of 9 criteria, namely working on questions, scoring guidelines, print quality, type and size of

letters with very good categories, as for layout, scoring objectivity, the practicality of instruments, time spent, and scientific literacy in reading with category 22 % very good and 78% good. For aspects of item validity consisting of 14 items, there are items with a good category of 79% and very good at 21%. From the validator's suggestion, the key answer to question number 4 needs to be improved, and the image layout improved.

Overall, the cognitive literacy-based scientific instruments developed can be used with little revision. After revision, the results are checked again by the validator. Then the revised cognitive instruments can be tested by the product.

Product trials were carried out in one class that had received an explanation of sound material, namely class VIII D of SMP Negeri 15 Banjarmasin. The number of students is 31 people. Furthermore, the scores obtained by students on the product trial will be analyzed to determine the validity, reliability, level of difficulty, and differentiation of the instrument.

The summary of the results of the analysis of the validity and reliability of the questions is as follows:

TABLE VI. ANALYSIS RESULT OF DIFFICULTY LEVEL

Criteria	No. Item
Difficult	3, 12
Medium	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14
Easy	-

Table 7 shows that in 14 items, there are 10 items that have moderate difficulty, 2 items that have difficult difficulty, and there are no items that have easy difficulty. As for the summary of the distinguishing index analysis of the product trials are as follows:

TABLE VII. CALCULATION OF DISTINGUISHING INDEX

Criteria	No. Item
Acceptable	2, 4, 6, 7, 8, 11, 12
Acceptable with revision	14
Revise needed	3, 10
Unacceptable	1, 5, 9, 13

Table 7 shows that in 14 items, there are 10 items that have success rates. Based on the table above, there are 4 items that must be discarded, namely items number 1, 5, 9, and 13. The questions number 2, 4, 6, 7, 8, 11, and 12 are acceptable, for items number 14 can be accepted and need to be corrected, while items number 3 and 10 must be corrected. Thus, from the results of product trials, 10 items were found that were valid and appropriate to be used for trial use. This is in accordance with the reference [12] if the data obtained is valid, then the instrument can be declared valid because it can state the description of the correct data in accordance with the truth.

The usage trials were carried out in three schools in Banjarmasin, namely Banjarmasin Public Middle School 24, Banjarmasin Public Middle School 4, and Banjarmasin Public Middle School 15 which were held from April to May 2017 and were attended by 203 Grade VIII Middle School students. The results of the usage trials are obtained as follows medium

level, 2 items that have a difficult level of difficulty, and there are no items that have easy difficulty. As for the summary of the power analysis different product trials are as follows:

TABLE VIII.   ANALYSIS RESULTS OF VALIDITY OF PROBLEMS

Outfit		Pt-Measure Corr	Note
MNSQ	*zSTD*		
1,17	1,3	0,70	Valid
1,26	2,4	0,43	Valid
1,22	1,3	0,55	Valid
1,20	1,9	0,57	Valid
1,00	0,0	0,58	Valid
0,98	-0,1	0,58	Valid
0,65	-1,7	0,63	Valid
0,88	-1,0	0,73	Valid
0,75	-1,9	0,74	Valid
0,58	-4,1	0,58	Valid

In Table 14. the above can be seen in the MNSQ outfit, zSTD outfit, and PT-Measure Correlation showing that all items are not outside the criteria simultaneously, then all items are appropriate and stated to have good validity. The questions tested can be said to be valid, even though there are some questions that have low and very low validity. Questions that are very low validity must be removed (no trial use is used) because the question is not sufficient to measure what is to be measured, namely measuring students' literacy skills. This is supported by reference [10] that a test must be valid, meaning that it can measure something to be measured. Although it contains valid items, it does not necessarily have high validity.

The Cronbach alpha value is 0.86, meaning that the reliability between the person and the items in question is overall very good. The average value of MNSQ output and ZSTD output for a person is 0.96 and 0.1 means that the quality is good, while the average value of MNSQ output and ZSTD output for items is 0.97 and -0.2, the quality of the grain is good.

There are four groups of items based on the level of difficulty, namely items that are very difficult, difficult, easy, and very easy. The percentage for items that are very difficult is 20%, items that are difficult are 30%, items that are easy are 30%, and items are very easy which is 20%. The questions that have the highest difficulty level are questions number 9 and 7 and the lowest which are number 3 and 1. The questions number 4 have the same level of difficulty as item number 5, and item number 6 has the same level of difficulty as item number 8.

As for the distinguishing analysis, there are 2 items that cannot be used, namely item 1 and item 2 because the two questions cannot distinguish students who have high abilities with students who have low abilities. This situation is caused by the problem number 2 has a high level of difficulty so that both students who have the ability above or below are both

many who cannot complete the number, while number 1 has a very difficult level of difficulty, making it difficult to distinguish students' abilities. Items cannot distinguish between high-ability students and low ability students because these items are too easy or too difficult.

IV. CONCLUSION

Based on the results of the development and trial, the conclusion is that the development of cognitive instruments based on scientific literacy on sound material developed can be declared valid, reliable, have varying degrees of difficulty, and good differentiation. This is supported by the following: 1) the cognitive instruments developed are valid based on the results of the Rasch program analysis (MNSQ outfit values, ZSTD outfit, and item measurement correlations), 2) the cognitive instruments developed are very reliable based on the program Rasch (Cronbach alpha with excellent category), difficulty level of cognitive instruments with Rasch program the difficulty level of questions varies, which is very difficult 20%, 30% difficult, 30% easy, and 20% very easy, and 4) the distinguishing power of cognitive instruments developed, 80% of the questions were accepted, and 20% of the questions were discarded.

REFERENCES

[1] M. Mustari, "Pengembangan instrumen ranah kognitif pada pokok bahasan fluida statis SMA/MA," Jurnal Ilmiah Pendidikan Fisika Al-Biruni, vol.5, no. 1, pp. 121–129, 2016.

[2] M. Muslich, Authentic assessment: penilaian berbasis kelas dan kompetensi, Bandung: PT Remaja Rosdakarya, 2011.

[3] U. Toharudin, Membangun literasi sains peserta didik, Humaniora, Bandung: Humaniora, 2011.

[4] Z. Zainab, M. Wati, and S. Miriam, "Pengembangan instrumen kognitif literasi sains pada pokok bahasan tekanan di kelas VIII SMP kota Banjarmasin," Jurnal Ilmiah Pendidikan Fisika, vol. 1, no.3, pp. 113–125, 2017.

[5] C. Pangestu, A. Abdurrahman, and F. Sesunan, Pengembangan buku saku IPA terpadu sebagai media pembelajaran berbasis literasi sains siswa SMP. Journal UNNES, vol. 43, no. 1, pp. 97–107, 2014.

[6] W. P. Astuti, A. P. B. Prasetyo, and E. S. Rahayu," Pengembangan instrumen asesmen autentik berbasis literasi sains pada materi sistem ekskresi," Journal UNNES, vol. 43, no.2, pp. 94–102, 2014.

[7] P. Setyosari, Metode penelitian pendidikan dan pengembangan, Jakarta: Kencana, 2013.

[8] E. P. Widoyoko, Evaluasi program pembelajaran, Jurnal Ilmu Pendidikan, Yogyakarta: Pustaka Belajar, 2016.

[9] R. Ratumanan, and L. Laurens, Penilaian hasil belajar pada tingkat satuan pendidikan. Ambon: Unesa University Press. 2011.

[10] B. Sumintono, and W. Widhiarso, Aplikasi pemodelan rasch pada assessment pendidikan, Cimahi: Trim Komunikata, 2015.

[11] A. Rusilowati, Pengembangan instrumen penilaian. Semarang: Unnes Press, 2014.

[12] S. Arikunto, Dasar-dasar Evaluasi Pendidikan Edisi 2, Jakarta: Bumi Aksara, 2013.