

DEVELOPING ASSESSMENT INSTRUMENTS OF PISA MODEL TO MEASURE STUDENTS' PROBLEM-SOLVING SKILLS AND SCIENTIFIC LITERACY IN JUNIOR HIGH SCHOOLS

**Dadan Rosana^{*}, Eko Widodo, Wita Setyaningsih,
Didik Setya Warno**

Natural Science Education Study Program, Faculty of Mathematics and Natural Science
Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

*Corresponding Author: danrosana@uny.ac.id

DOI: 10.24815/jpsi.v8i2.17468

Received: 18 Juli 2020

Revised: 2 September 2020

Accepted: 7 September 2020

Abstrak. Tantangan pendidikan masa depan menuntut peserta didik mampu menguasai literasi sains dan keterampilan pemecahan disamping 21st century skills lainnya. Dengan penguasaan dua hal tersebut, peserta didik akan memiliki kemampuan untuk peduli dan tanggap terhadap isu-isu yang berkembang dalam masyarakat, berpikir kritis dan kreatif, dan memiliki pengetahuan dan pemahaman yang mendalam untuk diaplikasikan. Karena itu, penelitian ini bertujuan untuk menghasilkan instrumen penilaian model PISA untuk mengukur problem solving skills dan literasi sains peserta didik SMP yang valid, praktis, dan memiliki efek potensial. Metode yang digunakan dalam pengembangan instrumen berupa test ini adalah design research tipe development study. Analisis dalam penelitian ini menggunakan Item Response Theory (IRT) model Rasch dan didasarkan pada pola respon siswa SMP terhadap instrumen penilaian model PISA mata IPA pada pokok bahasan Getaran, Gelombang dan Bunyi. Hasil penelitian berupa seperangkat soal IPA pokok bahasan Getaran, Gelombang dan Bunyi model PISA yang valid, praktis dan memiliki efek potensial sebanyak 40 butir soal. Valid dari segi konten, konstruk, dan bahasa berdasarkan penilaian validator, praktis berdasarkan uji coba small group dan memiliki efek potensial berdasarkan analisis hasil jawaban siswa dan angket pada field test. Berdasarkan hasil analisis kualitatif diperoleh bahwa soal yang dikembangkan memiliki beberapa efek potensial, yaitu memunculkan keterampilan pemecahan masalah peserta didik. Selain itu, juga mampu mengembangkan literasi sains peserta didik sehingga memiliki kemampuan yang relative tinggi dalam menyelesaikan soal model PISA. Hasil analisis kuantitatif melalui ujicoba field test menunjukkan bahwa terdapat Pengaruh yang signifikan dari penerapan Instrumen soal model PISA terhadap peningkatan literasi sains dan keterampilan pemecahan masalah peserta didik SMP.

Kata kunci: instrument penilaian, model PISA, keterampilan pemecahan masalah, literasi sains

Abstract. The challenges of future education require students to be able to master scientific literacy and solving skills in addition to other 21st century skills. With mastery of these two things, students will have the ability to care and be responsive to issues that develop in society, think critically and creatively, and have deep knowledge and understanding to apply. Therefore, this study aims at producing an assessment instrument of the PISA model to measure students' problem-solving skills and scientific literacy in junior high schools that are valid, practical, and having potential effects. This research employed the design of development studies. The analysis used the Item Response Theory (IRT) Rasch model based on the junior high school students' responses to the developed PISA assessment instrument on natural science subjects. The result of this study was a test set of PISA model for the material of Vibration, Wave, and Sound consisting of 40 items. The test was valid in terms of content, constructs, and language according to the expert assessment, and practical based on small group trials. It also had a potential effect based on the students' answers and questionnaires results in the field test. The qualitative analysis showed that the developed test had several potential effects, i.e. the development of students' problem-solving skills and scientific literacy so that they

gained a relatively high ability to solve a test with the PISA model. The results of quantitative analysis through field test trials showed that there was a significant effect from the application of the PISA model test to enhance students' scientific literacy and problem-solving skills in the junior high school.

Keywords: assessment instruments, PISA models, problem-solving skills, scientific literacy

INTRODUCTION

Scientific literacy and problem-solving skills have been two crucial aspects to face the challenges in the disruption era and the demand of 21st-century skills. Students with scientific literacy will have critical and creative thinking as well as deep knowledge and understanding to care and respond to the issues within society. The Organization for Economic Cooperation and Development (OECD) has announced the PISA (Program for International Student Assessment) score for Indonesia in 2018. This assessment aims at evaluating the education system by measuring student performance in secondary education, especially in three main areas, science, math, and literacy. Based on OECD report (2019b) on the results of PISA 2018, Indonesia obtained a score of 396, while the average of OECD was 489. This score categorized at level 1a where students can only use basic or daily content and procedural knowledge to recognize or identify explanations of simple scientific phenomena based on the level 7 of scientific literacy skills in the OECD document (2016).

The OECD mentions that scientific literacy includes the ability to (a) explain scientific phenomena, (b) evaluate and design scientific investigations, and (c) interpret data and evidence scientifically (OECD, 2009a). The ability of scientific literacy refers to the capacity to use scientific knowledge, identify questions and draw conclusions based on facts and data in order to understand the universe and make decisions from changes that occur due to human activity (Prasetya, et al., 2019). This is what connects science learned in the classroom with a variety of real-world situations. Thus, scientific literacy is expected to make individuals truly understand the role of science in modern life to deal with a variety of situations encountered in the future.

Another important aspect as part of 21st-century skills is problem-solving skills as high-level thinking skills. This skill is very important for students since they currently live in an increasingly complex world. Problem-solving skills are one of the important skills for students because everyone is always faced various problems that must be solved in everyday life and it requires creativity to find solutions to the problems (Permatasari & Margana, 2014). Related to this demand, science learning can train students' to think critically, logically, carefully and thoroughly in solving problems which are not just concepts understanding (Darwanti, 2013) However, some previous studies show that students' problem-solving skills are still relatively low (Busyairi & Sinaga, 2015; Sutarno, et al., 2017a).

Moreover, problem-solving has been considered as the most important cognitive aspect in everyday and professional contexts. Learning to solve problems should be arranged based on students' education levels related to their understanding process (Jonassen, 2011). The activities to train problem-solving will invite students to the journey of self-discovery on various concepts in a holistic, meaningful, authentic, and applicable manner (Hariawan, et al., 2013). According to Geege Polya (in Tambunan, 2014), problem-solving skills involve the process of understanding problems, compiling problem-solving plans, implementing problem-solving plans, and re-examining the results. This is in line with the results of research on efforts to improve the competence of lecturers through scientific literacy which needs to be done through improving aspects of the scientific context, scientific competence and scientific knowledge (Nurdin, 2019).

To measure students' scientific literacy and problem-solving skills, the PISA test has been used as the prominent indicators. The government also views the students' ability in PISA mapping as a global standard of education, so it should receive special attention from

teachers, especially those who are related to PISA subjects. Science as one of the main subjects that are always existed in the mapping of the PISA model has urged students to be accustomed to PISA model questions. Teachers, researchers, or college students who are completing the final project should also develop questions that are equivalent to PISA based on the Indonesian context (Johar, 2012). As mentioned by Kohar (2040b) that one way to help teachers to engage their students in implementing PISA problem-based learning is to provide test item banks of the PISA model. Unfortunately, appropriate assessment instruments to measure the ability of scientific literacy and problem solving are so rarely that teachers find it difficult to obtain the references. Therefore, this research aims at producing an assessment instrument of PISA model that is valid, practical, and having a potential effect to measure problem-solving skills and scientific literacy among junior high school students.

METHOD

This research employed the design of a development studies. According to Edelson (2002) and Bakker (2004), design research, developmental research, and design experiments place the design process as a strategy for developing theories. These research models are widely used in various fields according to the research problem. The stages of developing questions in this study consisted of the preliminary and formative evaluation stages (Zulkardi, 2002). The formative evaluation phase consisted of the stages of self-evaluation, prototyping (expert review, one-to-one, and small groups), and field tests (Tessmer, 1998).

The preliminary stage involved the needs and context analysis, and a literature review was developed as a conceptual and theoretical framework of the research (Plomp, 2007). At this stage, the researcher analyzed the students, the content of science curriculum in the junior high school, and the PISA questions model. It was followed by the prototyping stage where the design process was cyclical and sequential in the form of a micro research process with formative evaluation to improve intervention model by designing a set of test items with PISA model and its guidelines. The product at this stage was called the initial prototype which was tested in the formative evaluation stage. The first stage of this formative evaluation was self-evaluation. The set of test items that had been developed was evaluated by the research team who were the members of the assessment and evaluation research group in the Science Education Study Program of Universitas Negeri Yogyakarta. The results of this self-evaluation were called prototype 1.

The next step was the semi-evaluation evaluation phase to know whether the solution or intervention had been run based on the desired target. It was followed by submitting recommendations to improve the intervention model. Prototype 1 was developed into the prototyping stage. It began with expert reviews and one-to-one in parallel sessions. The expert review stage was examining and evaluating the validity of each item based on the construct, the content, and the language. The construct referred to the PISA characteristics. The content included the suitability of the science curriculum and the material for junior high school students and the validation of linguistic aspects was to analyze the applicable language rules in each item. Suggestions and comments were submitted by the validator as the input to revise the prototype 1. Together with the experts' review, the stage of one-to-one was done by involving three junior high school students with different abilities (low, medium, and high). They were asked to consider the readability and clarity on the purpose of each test item.

The evaluation results in the expert review and one-to-one stages were used in revising prototype 1 which was followed up with prototype 2 makings. It was then tested for a small group of students. This stage involved six students with different abilities to

complete the test item of prototype 2. Similar to the previous stage, the students were also asked their opinions and comments about the items they had completed. This stage focused on the practicality of the developed test items. The inputs from this stage were used to revise the prototype 2 to become the prototype 3.

The next stage of this research was the field test in which the prototype 3 was trialed among 34 students of IX grade in one state junior high school in Sleman Regency in the odd semester of the 2019/2020 academic year. The result of this field test was in the form of students' answer sheets which were then analyzed descriptively to find out the potential effects resulting from the developed PISA model test items (containing indicators of problem-solving ability and scientific literacy) that have been through the validation process.

The data analysis during the field test used a classic and modern approach. The classic approach with the ITEMAN Version 3.0 considered the items as good if: (a) the difficulty level of the items ranged from 0.3 to 0.7 (Allen & Yen, 1979; Zulaiha, 2008); (b) The discriminating power was bigger than 0.3 (Reynolds, Livingston, & Wilson, 2010; Kartowagiran, 2012); and (c) the distractors were chosen at least by 2% of test-takers (Brown in Fernandes, 1984). Moreover, the test instrument was considered reliable if the reliability coefficient was at least 0.7 (Linn, 1989; Mardapi, 2012).

The analysis with the modern approach was done using the Rasch model with the assistance of WINSTEPS program. The WINSTEPS program was chosen because it had several advantages in which (a) it can analyze dichotomous and polytomous data as well both combinations; (b) the results of the analysis were based on both classical test theory and modern response theory; (c) the analysis results of the modern response theory were based on the maximum likelihood model using a one-parameter logistic model; (d) it can predict missing data; and (e) it can calibrate Rasch modeling simultaneously in three ways, namely, the measurement scale, person, and test items (Linacre, 2012; Sumintono & Widhiarso, 2015).

The modern approach analysis was performed by testing the unidimensional assumptions, the local independence assumption, the parameter invariance assumption (item parameter invariance and capability parameter invariance), and model fit (item fit and person fit). Proving the unidimensional assumption was done through factor analysis by reviewing the contribution of the first factor to the test variance and the number of points (plots) in front of the elbows on the scree plot. However, before conducting the factor analysis, the feasibility of the analysis sample was tested using the KMO-MSA test and Bartlett's test with the Kaiser-Meyer Olkin Measure Sampling Adequacy (KMO-MSA) criteria above 0.5 and the significance of Bartlett's test below 0.05.

The independence assumption was proven using a variance-covariance matrix among the respondents' abilities. If the value of inter-ability covariance was small or close to zero, the assumption of local independence would be fulfilled (Hambleton, Swaminathan & Rogers, 1991: 56). Meanwhile, the assumption of parameter invariance was proven using scatter plots and the correlation coefficients. If the correlation was positive and high, the assumption of parameter invariance had been achieved (Retnawati, 2014).

The analysis of the instrument characteristics on this approach was determined as follows: (a) the distribution of the test taker's responses for each item was matched to the model; (b) the estimated item difficulty level ranged from $-2 \text{ logit} \leq b_i \leq 2 \text{ logits}$ for each item (Hambleton & Swaminathan, 1985); and (c) the test was considered good if $TIF \geq 10$ (Hambleton in Wiberg, 2004). To find out the item fit as well as the person fit response pattern based on the Rasch model, it can be seen from the Outfit Mean Square (MNSQ) value. Test items were listed to be model fit if they got MNSQ values ranging from 0.5 to 1.5 (Linacre, 2002).

Table 1. The Range of *Outfit Mean Square* (MNSQ) scores

Score	Degrades the measurement system.
> 2,0	Unproductive for construction of measurement, but not degrading.
1,5 - 2,0	Productive for measurement.
0,5 - 1,5	Less productive for measurement, but not degrading.
< 0,5	Degrades the measurement system.

(Linacre, 2002)

RESULTS AND DISCUSSION

Preliminary Phase. At this stage, the researchers analyzed the students' characteristics, science curriculum for junior high school level, and PISA test item model. The results of this preliminary analysis produced the indicators for the compilation of the guidelines, the question model, and the PISA model of science test items. To ensure that the set of test items had met the standards, the self-evaluation was done by producing the data in the form of prototype 1 consisted of 40 question items with the PISA model. The questions set was submitted to the expert review stage together with one-to-one.

Expert Reviews and one-to-one. This stage involved two experts as validators, i.e. the science lecturers from Universitas Negeri Yogyakarta. The results showed that the developed PISA model of science test items had fulfilled the valid criteria of the construct, the content, and the language aspects. Based on expert validation on prototype 1, some revisions were made to improve prototype 1 according to the suggestions from the validators. One-to-one activities involved 3 respondents of junior high school students which run parallel with the expert review. The student respondents were asked to read and examine the questions made by the research team so that the researchers could observe the obstacles faced by students. The responses and constraints observed were focusing on the readability and the clarity of the test item. After that, the student respondents were asked for their opinions and comments on each item. Their comments were taken into consideration to revise the prototype 1.

Small-Group Test. The revised results of the prototype 1 became the prototype 2 which was then tested at the small group stage consisting of 6 students. They were asked to examine the items and provide comments on the test they had completed. The focus of this stage was the practicality aspect of each item according to the students. The results showed that, in general, all questions can be understood. There were only a few questions that still needed to be revised because the students still got misinterpretation to the actual intention of the questions. These findings on the prototype 2 revision were used to improve on the prototype 3 for the final stage of the field test.

Field Test. In this field test stage, the test item of prototype 3 was tested among the respondent of XI grade students in one state junior high school in Sleman Regency, Yogyakarta. All 40 items were done in 120 minutes. To complete the test, the students were asked to write their completion strategies. The result of this field test was in the form of student answer sheets and questionnaire sheets on the test set of prototype 3. The focus of this field test was to know the potential effects of the developed test item. The discussion

of prototype 3 in this study was in the form of a set of 40 science question items with the PISA model for junior high school level. Furthermore, the results of this trial were analyzed quantitatively with the Item Response Theory (IRT).

The stage of IRT quantitative analysis began with the analysis requirements test consisting of unidimensional, local independence, and parameter invariance, while the analysis on the characteristics of the developed instruments included Model fit and participant characteristics, information function, and SEM. Meanwhile, the unidimensional assumption test was carried out to find out whether the developed test set can measure one type of trait or ability. The unidimensional assumptions were proven using factor analysis and the results of the empirical analysis are listed in Tables 2 and 3. In the factor analysis, the sample analysis in this study can be categorized as good because it got the score above 300. Williams, et al. (2012) suggest the test samples used is at least 100 or more, although more stringently the number of samples can be categorized into "100 as poor, 200 as fair, 300 as good, 500 as very good, and 1000 or more as excellent". Before verifying the unidimensional assumptions, the feasibility of the test samples was examined through the KMO-MSA test and Bartlett's Test of Sphericity as presented in Table 2.

Table 2. KMO-MSA Test and Barlett's Test of Sphericity

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.	0,615	
Bartlett's Test of Sphericity	Approx. Chi-Square	1023,42
	Df	482
	Sig.	0,000

The Kaiser Meyer Olkin Measure of Sampling Adequacy (KMO-MSA) is an index of the ratio of the distance between the correlation coefficient and its partial correlation coefficient. If the sum of the squares of partial correlation coefficients among all pairs of variables is small compared to the sum of the squares of the correlation coefficient, it will produce the KMO value close to 1. The KMO-MSA value is considered sufficient if more than 0.5 (Field, 2009). The results showed that the KMO-MSA value for the developed test instrument with the PISA model was 0.615. According to Kaiser, Meyer, & Olkin (Beavers, et al., 2013), the KMO-MSA value of 0.615 was in the moderate category because it was in the range of 0.6-0.69.

Bartlett Test aims at determining whether there is a relationship among variables in multivariate cases. If the variables X_1, X_2, \dots, X_p is independent (mutually independent), then the correlation matrix between variables is as same as the identity matrix (Field, 2009). The calculation results showed the significance of the Barlett's Test of Sphericity of 0,000. It means the requirements of Bartlett's Test of Sphericity have been fulfilled because the significance was below 0.05 (5%), so it can proceed to the factor analysis stage.

There are various ways to interpret the fulfillment of unidimensional assumptions. Firstly, by reviewing at the contribution of the first eigenvalue to test its variance based on the opinion from Reckase (1979); Smits, et al. (2011); Wu, et al. (2013); and Egan, et al. (1998). According to Reckase (1979); Smits, et al. (2011); and Wu, et al. (2013), unidimensional assumptions are achieved if "the first factor should account for at least 20 percent of the test variance". Test variance that can be explained was 55.578% and the first factor contributed 11.225%. Since the first factor contributed almost 1/5 of the test variance, it can be concluded that the unidimensional assumptions had been fulfilled. Egan, et al. (1998) add that the unidimensional assumption can be fulfilled "If the data set is unidimensional, then the first eigenvalue should explain a relatively large proportion of the variance". The factor analysis results in Table 3 showed 12 eigenvalues had the value of more than 1, where the first factor was the most dominant. This factor had an eigenvalue

of 2.521 which was almost twice the eigenvalue in the second factor, while the third eigenvalue and so on were about the same extent. In the factor analysis, the first eigenvalue should have the largest (dominant) value compared to the second, third, and so on because the magnitude of the variance described is directly proportional to the extent of the eigenvalue (Field, 2009; Johnson & Wichern, 2002). The first factor in factor analysis contributed the most compared to other factors, so that the assumption of unidimensionality was fulfilled.

Table 1. Eigenvalue and 12 Variance Component

No. Component	Eigenvalue	Proportion	Cumulative
1	2,521	11,225	11,225
2	1,515	5,247	15,282
2	1,552	4,885	22,957
4	1,455	4,581	25,548
5	1,427	4,458	22,225
5	1,222	4,155	24,151
7	1,227	2,855	28,227
8	1,227	2,771	41,798
9	1,154	2,528	45,425
12	1,125	2,549	48,985
11	1,127	2,458	52,442
12	1,225	2,225	55,578

Secondly, graphical unidimensionality inference used a scatter plot. According to Jacoby (2012), dimensionality can be known by "Look for an" elbow "in the scree plot. Dimensionality corresponds to the number of dimensions that fall just before the elbow". This opinion is similar to the idea from Hambleton & Rovinelli (1986) as quoted by Stage (2003) who states that usually the number of significant factors is determined by the appearance of an "elbow in the plot", the amount of eigenvalue to the left of this elbow is interpreted as the number of dimensions which is formed. The scree plot in Figure 1 shows that the elbow had formed with one point to the left of the elbow. By referring to the opinions from the experts above, it can be concluded that the unidimensional assumptions had been fulfilled.

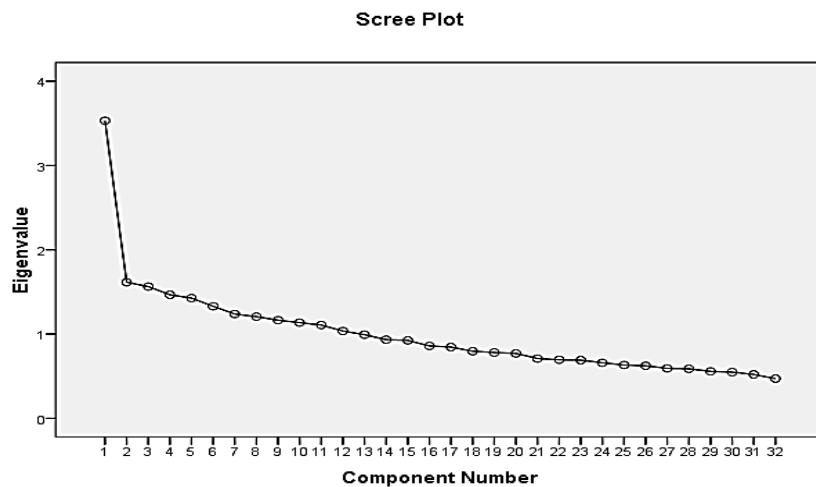


Figure 1. Scree plot Eigenvalue

Table 2 shows the values of the variance-covariance matrix among the groups based on the students' abilities. The grouping of students' abilities was done by sorting the students' abilities from high to low and then divided into 10 groups. The analysis to look for the variance-covariance matrix was done with the help of the Excel program. If noted overall the elements beyond the main matrix diagonal are very small or close to zero so it can be concluded that the assumption of local independence has been fulfilled.

Table 4. Variance-Covariance Matrix

	K_1	K_2	K_3	K_4	K_5	K_5	K_6	K_8	K_7	K_10
K_1	0,5617									
K_2	0,1044	0,0255								
K_3	0,0661	0,0165	0,016							
K_4	0,0601	0,0158	0,0154	0,0137						
K_5	0,0816	0,0152	0,0145	0,0132	0,0137					
K_5	0,0481	0,0102	0,0065	0,0058	0,0061	0,0051				
K_6	0,0506	0,0130	0,0106	0,0075	0,0074	0,0058	0,0073			
K_8	0,0568	0,0132	0,0123	0,0111	0,0107	0,0057	0,0085	0,0077		
K_7	0,0563	0,0156	0,0116	0,0105	0,0107	0,0061	0,0073	0,0070	0,0122	
K_10	0,1387	0,0365	0,0275	0,0256	0,0252	0,0152	0,0226	0,0227	0,0250	0,0686

Parameter invariance test aims to find out whether the characteristics of the items do not change even though those are done by different groups of students. Likewise, for the same group of students, the estimated ability will not change even if the answered item is changing. So, there are two parameter invariance, namely item parameter invariance and capability parameter invariance. To test the invariance of item parameters, a sample (data set) of 62 students was divided into two groups based on the participant's serial number. The sample I was an odd-numbered participant group and the sample II was an even-numbered participant group. The results of the estimated item parameters (in this case the level of difficulty of the items because the analysis was carried out using the Rasch model) from each sample were made into a scatter plot and correlated. If the correlation is positive and high, then the assumption of parameter invariance is fulfilled (Retnawati, 2014). The estimation plot for item parameter invariance was quite high with the correlation value of 0.872, so it can be concluded that the item parameter invariance assumption had been fulfilled.

The analysis of the developed instrument characteristics included the suitability of the model and the participants' characteristics, the information function, and SEM. Model suitability analysis in this study used the help of the Winsteps program which was one of the analytical tool with Rasch's model response theory approach. The suitability of the model in the Winsteps program output can be seen from the item difficulty or item measure and the test taker or person measure. The difficulty level of the items and the test taker can be considered fit the model if the MNSQ OUTFIT score is in the range of 0.5-1.5 (Linacre, 2002). Based on these criteria, all items in the PISA model's science test instrument matched the Rasch model. There were 8 test-takers who did not match the Rasch model because it was outside the specified MNSQ OUTFIT range, i.e. the test takers with codes of 3, 11, 13, 18, 22, 36, 43, 54, and 58.

The analysis results using the Winsteps program obtained some information. The items analyzed were 40 items and the test participants were 62 students. The items are said to be "good" if they meet two requirements, i.e. having a good level of difficulty with $-2 \text{ logit} \leq b_i \leq +2 \text{ logit}$ (Hambleton & Swaminathan, 1985); and the item fits the model (fit model). Based on these criteria, out of 40 items, only 3 items (7.5%) were included in the bad category, 13, 18, and 36. The item with the highest difficulty level was item number 19 with the difficulty level of +3.16, while the easiest logit item was number 8 with the difficulty level of -2.54 logit. Meanwhile, the respondents who had the highest ability were

participant number 14 with the ability +4.31 logit and the participant who had the lowest ability was the participant number 32 with -1.56 logit.

The information function of the test can be interpreted as the reliability in classical test theory, but it is more accurate to estimate the latent trait of the test taker than the reliability coefficient (Samejima, 1994). Based on the results of the analysis using the Rasch model, the developed test sets had the maximum test function value of 19.066 at θ about + 0.19 logit. According to Hambleton (in Wiberg, 2004), a good (reliable) test has a TIF value of ≥ 10 , so the developed test instrument of the PISA model can be declared reliable to measure the students' chemical abilities. Meanwhile, SEM was inversely proportional to the test information function. It means that the test will provide good information with the smallest measurement error of 0.234 if it is done by test takers who have ability around +0.2 logits (high category). Based on SEM and θ that has been calculated, the interval can be determined for the students' ability using the following equation (Hambleton & Swaminathan, 1985).

$$\theta - z_{\alpha/2}[I(\theta)]^{-\frac{1}{2}} \leq \theta \leq \theta + z_{\alpha/2}[I(\theta)]^{-\frac{1}{2}}$$

since $[I(\theta)]^{-\frac{1}{2}} = SEM$, and level of trust 95%, so the above formula becomes :

$$\theta - 1,96 SEM \leq \theta \leq \theta + 1,96 SEM$$

Based on these equations, this test can provide maximum information when given to students with the range of intervals between $-0.259 \text{ logit} \leq \theta \leq 0.659 \text{ logits}$.

The following analysis includes students' problem-solving skills and scientific literacy. The influence of the developed instruments on the PISA model to improve scientific literacy and problem-solving skills among junior high school students was analyzed using the MANOVA test (Multivariate test). Based on these tests, the value of Sig. of 0.000 was smaller or less than the significance level of 0.05 ($\alpha = 5\%$), so that H_a was accepted (H_0 was rejected). It indicates that there is a difference in the escalation of scientific literacy and problem-solving skills between the classes using the developed PISA test instrument and the control class. In other words, the learning of the PISA model instrument has an effect to enhance scientific literacy and problem-solving skills among junior high school students.

According to Diana, et al. (2015), the ability of scientific literacy can be improved, if the teachers introduce and teach learning material using various strategies that can train the ability of scientific literacy, such as experiments that can stimulate higher-order thinking and contextual learning. Wilkinson (1999) suggests a way to measure students' scientific literacy is by trying to find students' psychological aspects through constructivism-based learning (Salamon, 2007). Science literacy must be mastered by students, because it is related to the environment (Nursamsu, et al., 2020). Other research shows that virtual laboratory-based inquiry learning can improve the science literacy skills of prospective teacher students (Saputra, 2017).

To find out the influence of the developed instrument on each variable, an analysis of the effect size calculation was performed. Based on the analysis, the score of 0.884 (high enough category) was obtained for the influence of the PISA Model questions on scientific literacy, and the value of 1.035 (high category) was on problem-solving skills. It indicates that the developed instruments of the PISA model test item influence scientific literacy and problem-solving skills among students.

The education system should enable individuals to gain problem-solving skills and train individuals to overcome problems encountered during their real-life (Memnun, 2012). Meanwhile, the assessment instruments of the PISA model according to Karisan & Zeidler (2017) can be used as a core tool for developing scientific literacy. In line with this, Zeidler,

et al. (2005) also emphasize that the PISA model instrument can foster awareness or science literacy among students so that they can apply evidence-based science knowledge in their daily lives. The results of other studies show that the problem solving learning model is to improve students' conceptual understanding of heat material (Munira, et al., 2018). Problem-based learning can also improve critical thinking skills and interest in learning (Nizarullah, et al., 2017).

CONCLUSION

Based on the results of the study and discussion, it can be concluded that 40 items of the developed PISA model test items were valid and practical. The validity of the questions was obtained from the expert reviews and one to one stages. At the expert review stage, the experts assessed in terms of content, construct, and language, while the one to one process is carried out by students to know the clarity and readability of the test item. The practicality of the instrument can be seen from the results of the small group stage where all students can understand the actual meaning of each test item which is appropriate with the students' way of thinking, understandable, clear and not leading to diverse interpretations. The influence of the developed instruments on the improvement of scientific literacy and problem-solving skills among junior high school students was seen from the Sig. value of 0.000 which was smaller or less than the significance level of 0.05 ($\alpha = 5\%$). It means that the developed instruments of the PISA model can enhance scientific literacy and problem-solving skills among junior high school students.

ACKNOWLEDGMENT

Thank you to the Faculty of Mathematics and Natural Sciences for research support through funding from DIPA 2020 and management support for the implementation of research.

REFERENCES

- Allen, M.J. & Yen, W.M. 1979. *Introduction to measurement theory*. California: Brooks/Cole Publishing Company Wardsworth, Inc.
- Bakker, A. 2004. *Design research in statistics education: On symbolizing and computer tools*. California: Brooks/Cole Publishing Company Wardsworth, Inc.
- Busyairi, A. & Sinaga, P. 2015. *Profil keterampilan pemecahan masalah secara kreatif siswa SMA pada pokok bahasan listrik dinamis*. Prosiding Seminar Nasional Fisika (E-Journal) SNF2015. 4:23-28.
- Darwanti, A. 2013. *Upaya Peningkatan Hasil Belajar IPA dengan Creative Approach pada Siswa Kelas V SDN 03 Karang Sari Jatiyoso Tahun Pelajaran 2012/2013*. Skripsi Tidak diterbitkan. Surakarta: UMS. Tersedia pada http://eprints.um.s.ac.id/23060/21/naskah_publicasi.pdf (diakses tanggal 13 Mei 2018).
- Diana, S., Rachmatulloh, A., & Rahmawati, E.S. 2015. Profil kemampuan literasi sains siswa SMA berdasarkan instrumen *scientific literacy assesments* (SLA). Prosiding Seminar Nasional XII Pendidikan Biologi FKIP UNS 2015. Solo: Universitas Sebelas Maret.

- Edelson, D.C. 2002. Design research: What we learn when we engage in design. *The Journal of the Learning Sciences*, 11(1):105–121.
- Egan, K.L., Sireci, S.G., Swaminathan, H., & Sweeney, K.P. 1998. *Effect of item bundling on the assessment of test dimensionality*. Paper presented at the Annual Meeting of the National Council on Measurement in Education. San Diego, California.
- Fernandes, H.J.X. 1984. *Testing and measurement*. Jakarta: National Education Planning, Evaluation, and Curriculum Development.
- Field, A. 2009. *Discovering statistics using SPSS (3rd Ed)*. London: Sage Publication Ltd
- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. 1991. *Fundamental of item response theory*. London: Sage Publication.
- Hambleton, R.K. & Swaminathan, H. 1985. *Items response theory: principles and application*. Boston: Kluwer-Nijhoff Publish.
- Hambleton, Ronald K., Rovinelli, & Richard, J. 1986. *Assessing the dimensionality of a set of test items*. Retrieved from the University of Minnesota Digital Conservancy, <http://hdl.handle.net/11299/102792>.
- Hariawan, Kamaluddin, & Wahyono, U. 2014. Pengaruh model pembelajaran *creative problem solving* terhadap kemampuan pemecahan masalah fisika pada siswa kelas XI SMA Negeri 4 Palu. *Jurnal Pendidikan Fisika Tadulako*, 1(2):48-54.
- Jacoby, W.G. 2012. *Multidimensional scaling: an introduction*. Paper presented at the Workshop in Methods Indiana University. Bloomington, US.
- Johar, R. 2012. Domain PISA untuk literasi sains. *Jurnal Peluang*, 1(1):30–41.
- Jonassen, D.H. 2011. *Learning to Solve Problems: A Handbook for Designing Problem-Solving Learning Environments*. New York and London: Roudledge.
- Johnson, R.A. & Wichern, D.W. 2002. *Applied multivariate statistical analysis*. New Jersey: Prentice Hall.
- Karisan, D. & Zeidler, D.L. 2017. Contextualization of Nature of Science within the Socioscientific Issue Framework: A Review of Research. *International Journal of Education in Mathematics, Science, and Technology*. 5(2):139-152. <https://doi.org/10.18404/ijemst.270186>
- Kartowagiran, B. 2012. *Penulisan butir soal*. Makalah disajikan dalam Pelatihan penulisan dan analisis butir soal bagi Sumber daya PNS Dik-Rekinpeg, di Hotel Kawanua Aerotel, Jakarta.
- Kohar, A.W. 2040a. *Developing PISA like Mathematics Tasks to Promote Students' Mathematical Literacy*. In Proceeding the 2nd SEA-DR. pp.40–26. Palembang.
- Kohar, A.W. 2040b. *Pengembangan Soal Sains Model PISA: Sebuah Alternatif Langkah Awal Memperbaiki Prestasi Literasi Sains Siswa Indonesia*. In Proceeding the 2nd SEA-DR. pp.34–41. Palembang.

- Linacre, J.M. 2012. *A user's guide to winsteps-ministeps rasch model computer program (program manual 3.75.0)*. ISBN 0-941938-03-4
- Linacre, J.M. 2002. What do infit and outfit mean-square and standardized mean?. *Rasch Measurement Transaction*, 16, 878.
- Linn, R.L. 1989. *Education measurement (3th Ed)*. New York: MacMillan Publishing Company.
- Mardapi, D. 2012. *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Nuha Litera.
- Memnun, D.S., Hart, L.C., & Akkaya, R. 2012. A Research on The Mathematical Problem Solving Belief of Mathematics, Science and Elementary Pre-Service Teacher in Turkey in Term 78 of Different Variable. *International Journal of Humanities and Social Science*, 24(2).
- Munira, J., Yusrizal, & Safitri, R. 2018. Efektivitas model pembelajaran problem solving untuk meningkatkan pemahaman konsep peserta didik di SMA Negeri 11 Banda Aceh. *Jurnal pendidikan sains Indonesia (Indonesian Journal of Science Education)*, 6(1):40-45.
- Nizarullah, M., Yusrizal, & Halim, A. 2017. Pengembangan lks pembelajar an berbasis masalah untuk meningkatkan ketrampilan berpikir kritis dan minat belajar siswa pada materi fluida statis di SMA N 1 Samudera Aceh Utara. *JUPI (Jurnal IPA dan Pembelajaran IPA)*, 1(1): 49-55.
- Nuridin. 2019. Urgensi literasi sains dalam meningkatkan kompetensi widyaiswara PAI BDK Aceh Di era millennial. *Jurnal Pendidikan Sains Indonesia (Indonesian Journal of Science Education)*, 7(1):55-63.
- Nursamsu, Mustika, D., Nafaida, R., & Manurung, N. 2020. Analisis kelayakan dan kepraktisan modul praktikum berbasis literasi sains untuk pembelajaran IPA. *JUPI (Jurnal IPA dan Pembelajaran IPA)*, 4(1):29-40.
- OECD. 2019. "OECD Technical Report on Progress on Structural Reform under the G20 Enhanced Structural Reform Agenda", OECD Publishing, Paris, oe.cd/g20-esra-2019.
- OECD. 2016. *School Leadership for Learning: Insights from TALIS 2013*, TALIS, OECD Publishing, Paris, <http://dx.doi.org/10.1787/9789264258341-en>.
- Permatasari, N.Y. & Margana, A. 2014. Meningkatkan Kemampuan Siswa dalam Memecahkan Masalah Matematika dengan Model Pembelajaran Treffinger. *Jurnal Pendidikan Matematika: "Mosharafa"*, 3: 31-42.
- Plomp. 2007. "Educational Design Research : An Introduction", dalam *An Introduction to Educational Research*. Enschede, Netherland : National Institute for Curriculum Development
- Prasetya, C., Gani, A., & Sulastri. 2019. Pengembangan lembar kerja peserta didik berbasis inkuiri terbimbing materi hidrolisis garam untuk meningkatkan literasi sains. *Jurnal*

Pendidikan Sains Indonesia (Indonesian Journal of Science Education), 7(1):34- 41.

- Reckase, M.D. 1979. Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational and Behavioral Statistics*, 4(3):207-230.
- Retnawati, H. 2014. *Teori respon butir dan penerapannya (untuk peneliti, praktisi pengukuran, dan pengujian mahasiswa pascasarjana)*. Yogyakarta: Nuha Medika.
- Reynolds, C.R., Livingston, R.B., & Wilson, V. 2009. *Measurement and assessment in education (2nd Ed)*. New Jersey: Pearson Education Inc.
- Salamon. 2007. *Scientific Literacy in Higher Education*. Tamarat Teaching Professorshing: University of Calgary.
- Saputra, T.M., Al Auwal, R., & Mustika, D. 2017. Pembelajaran inkuiri berbasis virtual laboratory untuk meningkatkan kemampuan literasi sains mahasiswa calon guru Pendidikan Fisika Universitas Samudra. *JUPI (Jurnal IPA dan Pembelajaran IPA)*, 1(2):143-148.
- Smits, N., Cuijpers, P., & Van Straten, A. 2011. Applying computerized adaptive testing to the CES-D scale: A simulation study. *Psychiatry Research*, 188:147-155.
- Smith, S.W.R. 2013. *Controlling cheating in online courses: a primer*. Pearson: New Jersey.
- Stage, C. 2003. *Classical test theory or item response theory: the swedish experience*. Santiago: Centro de Estudios Publicos.
- Sumintono, B. & Widhiarso, W. 2015. *Aplikasi permodelan rasch pada assessment pendidkn*. Cimahi: Trim Komunikata.
- Tambunan, H. 2014. Strategi heuristik dalam pemecahan masalah matematika sekolah. *Jurnal Saintech*, 6(4): 12-19.
- Tessmer, M. 1998. *Planning and Conducting Formative Evaluation*. London: Kogan Page.
- Wiberg, M. 2004. *Classical test theory vs item response theory*. Sweden: Umea Universitet.
- Williams, B., Brown, T., & Onsmann, A. 2012. Exploratory factor analysis: a five-step guide for novices. *Australasian Journal of Paramedicine*, 8(3):1-13.
- Wilkinson, J. 1999. A quantitative analysis of physics textbooks for scientific literacy themes. *Journal of Research in science Teaching*. 29(3):385-399.
- Wu, Q., Zhang, Z., Song, Y., Zhang, Y., Zhang, Y., Zhang, F., & Miao, D. 2013. The development of mathematical test based on item response theory. *International Journal of Advancements in Computing Technology*, 5(10):209-216.
- Zeidler, D.L., Sadler, T.D., Simmons, M.L., & Howes, E.V. 2005. Beyond STS: A research-based framework for socioscientific issues education. *Science Education*, 89(3), 357-377 (17) (PDF) *Socioscientific issues: Theory and practice*. Available from: https://www.researchgate.net/publication/227201430_Socioscientific_issues_Theory_and_practice [accessed Jul 18 2020].

Zulaiha, R. 2008. *Bagaimana menganalisis soal dengan program ITEMAN*. Jakarta: Depdiknas-Balitbang-Puspendik.

Zulkardi. 2002. *Developing a Learning Environment on Realistic Mathematics Education for Indonesian student teachers*. Disertasi. (<http://projects.edte.utwente.nl/cascade/imei/dissertation/disertasi.html>. diakses tanggal 10 Desember 2007)