Development of a Computer-Based Diagnostic Assessment to Measure Junior High School Students' Readiness to Participate AKM and PISA with Automatic Feedback

Didik Setyawarno[1*], Dadan Rosana[2], Eko Widodo[3], Maryati[4]

[1,2,3,4]*Department of Science Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta, Indonesia.*
E-mail: didiksetyawarno@uny.ac.id

Abstracts: This study aims to develop a Computer-based Diagnostic assessment to measure the student's readiness to participate in AKM and PISA. The study used a Research and Development model with 4-D model consisted of four main stages, namely define, design, develop, and disseminate. Preliminary research was conducted by depth-analyzing the AKM material and science material in the science curriculum of junior high school, identifying science materials or concepts that developed as AKM and PISA assessment instruments, designing instrument models and Computer-based Diagnostic assessments, implementation or empirical tests, validation, and first revision. And it ended with the dissemination stage through international seminars, attended by students, teachers, lecturers, and observers of science education. The conclusion of the study shows that the developmental research of a 4-D model is feasible, valid, and effective to be used in developing computer-based diagnostic assessment products to measure the student's readiness of junior high school to participate in AKM and PISA with automatic feedback. Moreover, the conclusion of this research is that application of a computer-based diagnostic assessment is formed to measure the student's readiness of junior high school to participate in AKM and PISA with valid automatic feedback based on the Aiken formula and appropriate, according to science education experts from the aspect of content, construct, and language. The results of the empirical test of the application of computer-based diagnostic assessment are used to measure the student's readiness of junior high schoo to participate in AKM and PISA with automatic feedback stating that all items fit the Rasch model. And, the results of a limited-scale test for the student's readiness to participate in AKM and PISA showed that SMP N 4 Pakem obtained 67.92% readiness in the average category, while SMP N 1 Tempel obtained 48.75% readiness in the below average category.

Keywords: Development, Computer-based diagnostic assessment, Students' readiness, AKM, PISA, Automatic feedback.

1. INTRODUCTION

One of the problems in education in Indonesia that is quite prominent is the low score on the PISA test among junior high school students. PISA (Program for International Student Assessment) is an international assessment system that focuses on the ability of 15-years-old children in the fields of reading literacy, mathematical literacy, and scientific literacy. [1] The PISA assessment conducted by the OECD (Organization for Economic Co-operation and Development) is caused by the reason that literacy skills are very much needed in facing the challenges of the 21st century.[2] According to the results of the PISA assessment, Indonesian students aged 15 years who did not reach the OECD achievement standards in mathematics and science literacy were 76% and 67% [1]; [3]. Based on these results, the OECD recommends improving the language, mathematics, and science literacy among Indonesian students when these students are prepared to continue to high school or equivalent. While, for those who do not continue to high school, this achievement indicates a low level of performance to be able to adapt to a world full of change and competition (OECD, 2015). [4]

To answer the result of the PISA assessment, the Indonesian Ministry of Education and Culture issued an education assessment policy in 2021, namely the Minimum Competency Assessment (in Indonesian it is abbreviated as AKM) which is regulated in Permendikbudristek Number 17 of 2021. The AKM is one of the National Assessment Instruments (AN) as a substitute for the National Examination, where AKM measures two basic competencies, reading literacy and mathematical literacy. Reading literacy is defined as the ability to understand, use, evaluate, and reflect on various types of written texts to develop individual capacities as Indonesian citizens and global citizens and can contribute productively to society. [5] Numerical literacy is the ability to think using concepts, procedures, facts, and mathematical tools to solve everyday problems in various types of contexts that are relevant for individuals, as citizens of Indonesia and the world. The two literacy cover competencies (1) logical-systematic thinking skills; (2) Reasoning skills using the concepts and knowledge; (3) Skills to sort and process information[6]; [7].

The emergence of this new policy stimulates the interest of researchers, especially educational researchers to develop an assessment model that leads to AKM and PISA aiming to improve the quality of education in Indonesia. Along with the rapid use of computer technology, the assessment system has begun to shift from paper-based assessments to computer-based assessments.[8] So far, paper-based assessment is considered to have several weaknesses, such as it might not be carried out quickly (real-time) so the students might not know the test results directly and quickly. The weakness made the understanding of students' concepts might not be detected immediately. Also, the teacher has difficulty correcting answers and providing feedback to each student.[9]

Learning assessment is an important part of learning that cannot be abandoned. Traditionally, assessment is intended to find out and report the things that have been learned and their relation to class activities. Assessment is an integral part of teaching and learning activities in schools and mediates interactions between teachers and students in grades.[10] The current conditions amid the Covid-19 pandemic strongly emphasize the need to improve the quality of assessment in science learning in schools. The pandemic has brought several new policies in the field of education. For example, the government enforces limited learning with various patterns and ways, including hours of learning in face-to-face/offline learning, implementing online learning, or a combination of online and offline for elementary school, middle school, and college students.[11]

Students during the Covid-19 pandemic with limited learning modes desperately need a supporting assessment to measure learning outcomes, especially in the government programs of the AKM (minimum competency assessment). Assessments are carried out in various learning conditions, at the beginning of learning, during learning, or after learning. A form of assessment to improve the quality of learning is a diagnostic assessment.[12] A diagnostic assessment is used to find out the weaknesses of students in mastering certain materials or competencies and the factors caused. Recently, cognitive diagnostic assessments designed to measure the specific knowledge structure and process skills of learners have attracted a lot of attention.[13] The results of the diagnostic assessment might be used as a basis for follow-up with appropriate treatment (interventions) following the characteristic of students.

The follow-up of this diagnostic assessment is adjusted to the aspects assessed in the assessment. Follow-up of the learning reflects actions that are relevant to the conditions of each student, accommodating as well as flexible. Diagnostic assessment tests have different characteristics from traditional tests that only make a list of students. Diagnostic assessments can analyze multidimensional cognitive processes comprehensively toward the understanding of the material being studied and enable predictions of mastery or non-mastery of discrete sub-skills or attributes of learners involved in the process.[14] The questions in a Selected Response are reasoned. Then, it detects students' learning difficulties and not to test students "Pass" or "Fail". Teachers are expected to use the results of the diagnostic assessment to improve their teaching and make a good instruction plan to facilitate the development of students' competencies.

The type of assessment must be adjusted to the objectives of the assessment itself, especially in developing student competencies. Lately, there is an assessment that is widely used in the world of education, namely a diagnostic assessment. This assessment is nothing new. However, with the emergence of various obstacles in the world of education due to the COVID-19 pandemic, this assessment is starting to become popular again. The development of student competencies in the current era is very important for all educational institutions, especially for demanding competencies in the digital and global era. Several factors that require a diagnostic assessment are an international level assessment, namely the Program for International Student Assessment (PISA) and a government program of the minimum competency assessment (AKM).

One of the international level assessments is carried out by the International Student Assessment (PISA) and Indonesia is a member country that participates in the assessment regularly. The results of the Program for International Student Assessment (PISA) for Indonesia in 2018 show the score under the expected average score. In the three-year PISA program in 2018, Indonesia was ranked 74th out of 79 countries that participated in this program. The results showed that the ability of Indonesian students to read, achieved an average score of 371, with an OECD average score of 487. Then, the average score for mathematics is 379 with an OECD average score of 487. Furthermore, in science, the average score of Indonesian students is 389 with an OECD average score of 489.

Indonesia's position and score in the year 2018 showed declining results compared to the previous round (2015, 2012, 2009). But, Indonesia's ranking did not move from the bottom 10 [4]; [3].

These results indicate the low literacy among students in Indonesia, seen in the three abilities tested because PISA does not only measure knowledge but also how literate students are. Assessment and Science learning should facilitate students to achieve these abilities and skills. Also, it needs an appropriate assessment instrument. Appropriate assessment instruments that use to measure scientific literacy skills are still very rare. Here, it becomes one of the obstacles for teachers. Minimum Competency Assessment (AKM) is a government program besides the international assessment program (PISA) which requires a diagnostic assessment. AKM is defined as an assessment of basic competencies needed by all students to develop their capacity and participate positively in society. [5] AKM is used to measure students' cognitive abilities where the measured aspects are reading literacy skills and numeracy literacy,[15] including the competencies of logical-systematic thinking skills, reasoning skills using concepts and knowledge, and skill to sort and process the information. AKM presents problems with various contexts that students are expected to solve using their reading and numeracy literacy competencies. AKM is intended to depth-measure competence, not just content mastery.

The PISA and AKM encourage the implementation of innovative learning and assessment systems that are oriented to the development of higher-order thinking skills, namely reasoning skills, and not focus on memorization. The ability of teachers to develop international standard assessments of PISA is still varied and needs to be improved. [16] Moreover, there are still many teachers and prospective teachers who do not know about the concept of national assessment, especially the concept of Minimum Competency Assessment (AKM) which is used to measure students' cognitive abilities (Asrijanty, 2020); [15]. Then, based on these conditions, this program aims to provide insight into the concept, including member countries that participate in the assessment on a regular term. The results of the Program for International Student Assessment (PISA) and Minimum Competency Assessment (AKM), are specially designed for the development of the PISA and AKM model diagnostic assessments for science teachers.

Based on the facts and problems, the study tries to produce an assessment system that supports the achievement of AKM and PISA through R & D research which aims to develop a computer-based assessment system, namely Computer-based Diagnostic assessment through R & D research. At the same time, it is hoped that it become an effective and efficient assessment tool in the era of the COVID-19 pandemic, by reducing gatherings of people during the paper-pencil test and might be carried out flexibly by teachers and students with appropriate feedback.

2. RESEARCH METHOD

2.1. Research Design

The research was a Research and Development (R & D) using 4-D (Four D) model. The stages of development include define (limitation), design, develop, and disseminate [17]; [18]. The research is divided into four stages. The first stage is the define stage. The results of the define stage are the depth level of science material in junior high school in PISA and AKM. The second stage is the design that will be developed on campus. Next, the develop stage is conducted. In this stage, it produces instruments and computer-based diagnostic assessment models that have been developed and revised based on input from experts and empirical tests in several junior high schools. And, the disseminate stage of the results was conducted through national science education seminars aiming at the instruments that can be used on a wider scale by educational institutions.

The first stage is definition. The purpose of this stage is to determine and define the science materials in junior high school that is included in the AKM and PISA. It begins with an analysis of the objectives and limitations of the material developed by the assessment instrument using a computer-based diagnostic assessment. The stage includes the analysis of core competencies and basic competencies, concept analysis, and the formulation of AKM and PISA assessment indicators to measure scientific literacy, numerical literacy, and language literacy. The second is design stage. This stage aims to prepare a prototype of the developed diagnostic assessment computer-based device. This stage consists of preparing the AKM and PISA assessment instruments and selecting the appropriate

format. The instruments are prepared based on the results of the formulation of the AKM and PISA assessment objectives. The format selection stage is carried out by reviewing the existing formats and those that will be developed in the research.

The third is development stage. This stage aims to produce a computer-based diagnostic assessment that has been revised based on input from experts. This stage includes (a) drafting, (b) validation by experts, and (c) limited test with real students. The results of stages (b) and (c) are used as the basis for revision. The last one is dissemination stage. This stage is the stage of introducing to a wider scale which is presented in national and international seminars in the field of education, which are attended by various groups, such as teachers, lecturers, and students of science education.

2.2. Research Sites and Subjects

The research sites used in the empirical test were two junior high schools (SMP) in the DIY (Special Region of Yogyakarta), namely SMP N 1 Tempel and SMP N 4 Pakem. The subjects were science education lecturers, school science teachers, and junior high school students.

2.3. Data Analysis

The data were analyzed quantitatively and qualitatively. Qualitative data were obtained from input or comments from science education experts. It is used as the basis for the first revisions before the instruments are applied and used in schools. Moreover, quantitative data were obtained from the expert assessment sheet to see the quality of the instrument items that developed from the content, construct, and language aspects to assess the feasibility of the computer-based diagnostic assessment instrument, aiming to measure the readiness of students in AKM and PISA. Field tests were conducted to see the quality of items empirically and applied in junior high schools in the DIY. The following are details of the data analysis.

a. Analysis of validation in terms of content, constructs, and language from the results of expert judgment for each item by an expert was analyzed using the Aiken formula [19]; [20].

$$V = \frac{\sum S}{[n(C-1)]} \qquad (1)$$

Where: $S = R - l_o$

Description:

Lo = The lowest rating scorr (eg 1)

C = Highest rating score (eg 4)

R = Score given by raters/validators

N = Number of raters or validators

b. Scale conversion analysis of the quantitative to qualitative from the results of expert assessments related to the feasibility of the AKM and PISA assessment instruments in computer-based diagnostic assessments that have been developed from the content, construct, and language aspects with the following criteria[16]; [19].

Table 1. Score Conversion Scale to Criteria.

Expert Assessment Score	Criteria
X >Xi + 1,8 Sbi	Very Good
Xi + 0,6 SBi < X ≤ Xi + 1,8 SBi	Good
Xi - 0,6 SBi < X ≤ Xi + 0,6 SBi	Average
Xi – 1,8 SBi < X ≤ Xi – 0,6 SBi	Poor
X ≤Xi - 1,8 Sbi	Very Poor

c. Analysis of fit-item with the Rasch model (model 1 PL) used the QUEST application. The fit-items with the model in the QUEST program are based on the average value of INFIT Mean of Square (INFIT MNSQ) and its standard deviation or the mean of INFIT Mean of Square (INFIT MNSQ) or INFIT t. The fit of each item with the model in the QUEST program is based on the score of the INFIT MNSQ value or the INFIT value of t items with the following conditions in table 3 [21]; [20]; [22].

Table 2. Infit MNSQ on the Rasch Model.

INFIT MNSQ	Where
> 1,30	Not fit with Rasch model
0,77 – 1,30	Fit with Rasch model
< 0,77	Not fit with Rasch model

d. Descriptive statistical analysis was conducted using the Ms. application. Excel to see an overview of each class from the AKM and PISA assessments, including the mean value, standard deviation, variance, maximum value, and minimum score [23].

e. Analysis of student's readiness from the results of the trial was carried out with the percentage, presented in Table 3 [24]–[26].

Table 3. Categories of Students' Readiness to Participate in AKM and PISA.

Percentage (%)	Category
86 – 100	Very Good
76 – 85	Good
60 – 75	Fair
55 – 59	Poor
≤54	Very Poor

3. RESULT AND DISCSSION

The type of research was Research and Development (R & D) which is applied to the evaluation of science learning to produce computer-based diagnostic assessment products to measure the students' readiness of the junior high school to participate in AKM and PISA with automatic feedback. The development model used 4-D which consists of four main stages, namely define, design, develop, and disseminate[17]; [18]. The model is one of the appropriate models for the field of Education. Research with a 4-D model is divided into four stages. The first stage is the define stage. And, it obtained a diagnostic assessment instrument guideline following the science curriculum at the junior high school level. The second stage is the design of a computer-based assessment that will be developed for science subjects. The third stage is the preparation and development of an automatic computer-based diagnostic assessment which involves experts and field tests to produce products that are fit to be used to measure the student's readiness

of the junior high school to participate in AKM and PISA with feedback. The fourth stage is the dissemination of results through science teacher training activities at SMP Sleman-Yogyakarta. In the end, the results can be adapted for the evaluation of science learning in schools.

3.1. Define Stage

The purpose of this stage is to determine and define a diagnostic assessment instrument that fits with the science curriculum at the junior high school level. This stage is carried out through literature studies and previous research, including analysis of core competencies and basic competencies, and analysis of materials applied in the curriculum 2013 for science subjects. The results of the analysis at this stage are in the form of a diagnostic assessment guideline that will be developed and related to AKM and PISA[1]. AKM has three assessment components, namely: content, cognitive processes, and context. While PISA has four aspects of student performance related to scientific literacy, namely context, knowledge, competence, and attitudes[7]. Based on the components or domains of AKM and PISA, there is an integration between AKM and PISA in the context of scientific literacy, including knowledge in PISA which is closely related to content and cognitive processes in AKM. And, both are closely related to contexts or cases relevant to science[27]; [28]. The diagnostic assessment instrument guideline to measure the dimensions of the integration of AKM and PISA is presented in Table 4.

Table 4. Integration of AKM and PISA and Assessed Competency Indicators.

Integration of AKM & PISA	Scientific Competence in Scientific Literacy	Indicators used in the preparation of the diagnostic assessment
Context, Content, Knowledge, and Cognitive Processes	Explaining phenomena scientifically	a. Remembering and applying scientific knowledge. b. Identify, use, and create simple picture models to explain scientific phenomena encountered in everyday life. c. Make appropriate predictions. d. Offer an explanatory hypothesis. e. Describe the potential engagement of scientific knowledge for society.
	Evaluating and designing scientific investigations	a. Identify questions that investigated in scientific studies b. Propose a way of investigating a particular question scientifically (problem formulation) c. Evaluating a method of scientific inquiry d. Evaluating the methods used by scientists to ensure data reliability and objectivity of an explanation.
	Interpreting scientific evidence and data	a. Converting data from one form to another (diagrams, graphs, etc.). b. Analyzing and interpreting data to draw appropriate conclusions. c. Identifying assumptions, evidence, and reasons related to science. d. Distinguishing arguments based on evidence and scientific theory from arguments based on judgment. e. Evaluating scientific arguments and evidence from various sources

3.2. Design Stage

This stage is to design a prototype on a digital platform by using a google form that is set to a computer-based diagnostic assessment to measure the student's readiness for junior high school to participate in AKM and PISA with automatic feedback. The format selection stage is carried out by reviewing the existing formats and those that might choose to be developed in the research. The product design is an online computer-based application using a google form, consisting of questions with the AKM and PISA models. The questions are multiple-choice and correct or wrong statements. This form was chosen to facilitate implementation and data collection. In addition, Google Forms can provide feedback features automatically in the form of writing or video, or audio-visual. The development design is presented in Figure 1.

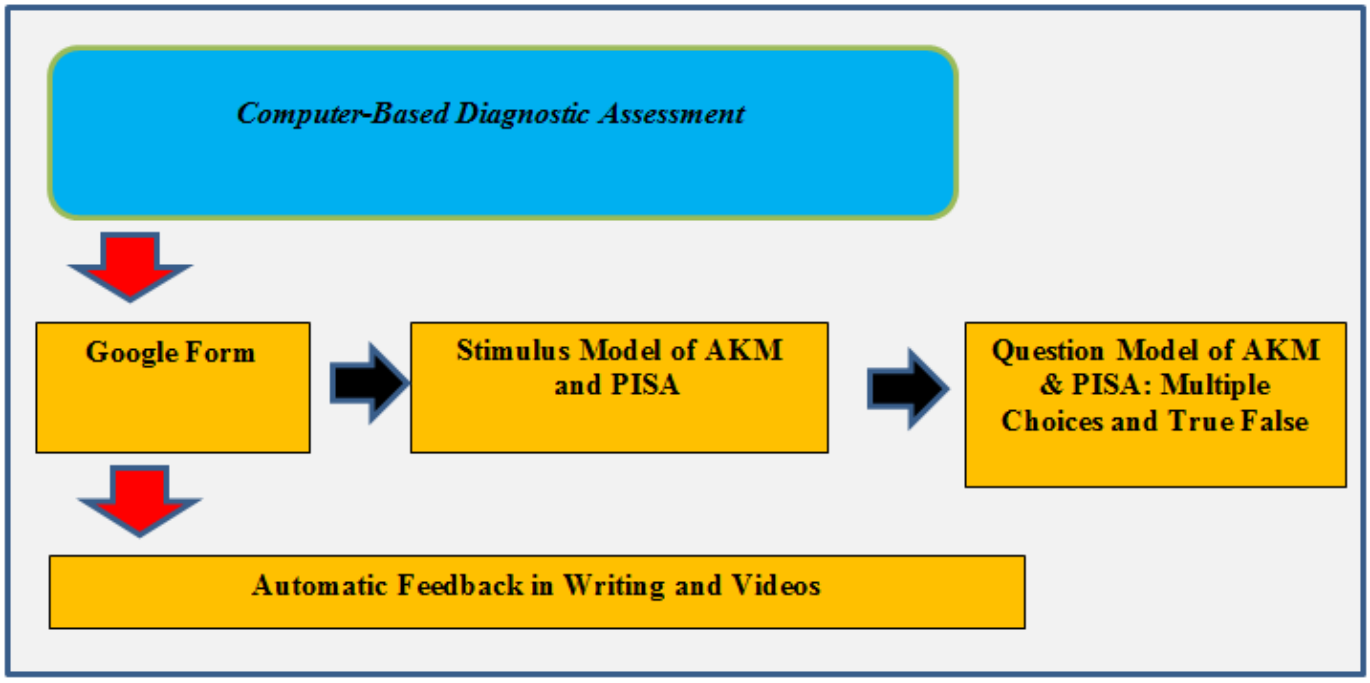


Figure 1. Development Design of Computer-Based Diagnostic Assessment Application.

Moreover, this stage consists of a guideline of diagnostic assessment instruments in science subjects that measure the AKM and PISA dimensions, consisting of the aspects of content, context, knowledge, and cognitive processes in the context of scientific literacy. Instruments are arranged based on the results of the formulation of the previously identified instrument guideline. The products of the planning stage are indicators of diagnostic assessment instruments, initial drafts of questions, and discussions to measure the student's readiness for junior high school to participate in AKM and PISA is presented in Table 5.

Table 5. Indicators of Diagnostic Assessment Instruments to Measure the Student's Readiness of Junior High School to Participate in AKM and PISA.

Stimulus Question (context)	Content and Knowledge	Scientific Literacy	Cognitive Level	No of questions
Drawing of weighing process	Measurement	Evaluating and designing scientific investigations	C4	1
		Evaluating and designing scientific investigations	C4	2
Figures and graphs of the density of air in the area	Heat	Explaining the phenomenon scientifically	C4	3
			C4	4
Figures and tables on heat transfer	Heat.	Interpreting scientific evidence and data	C4	5
		Explaining phenomena scientifically	C4	6
		Evaluating and designing scientific investigations	C4	7

Figure of the photosynthesis process	Energy and photosynthesis in plants	Evaluating and designing scientific investigations	C4	8
		Explaining phenomena scientifically	C4	9
Figures and narratives of tsunami phenomena and plate tectonics	Earth's structure and natural phenomena	Explaining phenomena scientifically	C4	10
		Evaluating and designing scientific investigations	C4	11
		Explaining phenomena scientifically	C4	12
		Interpreting scientific evidence and data	C4	13
		Interpreting scientific evidence and data	C4	14
Figures and Comparative data of terrestrial and jovian planets and graphs of the speed of revolution of the planets and the distances of the planets to the sun	Solar system	Evaluating and designing scientific investigations	C4	15
		Interpreting scientific evidence and data	C4	16
		Evaluating and designing scientific investigations	C4	17
Figure of Free fall motion experiment and case pictures of flying flies	Motion and force	Evaluating and designing scientific investigations	C4	18
		Interpreting scientific evidence and data	C4	19
		Interpreting scientific evidence and data	C4	20
Figures that have similarities and table of the living things classification	Classification of living things	Interpreting scientific evidence and data	C4	21
		Interpreting scientific evidence and data	C4	22
		Evaluating and designing scientific investigations	C4	23

		Evaluating and designing scientific investigations	C4	24
		Interpreting scientific evidence and data	C4	25
Figures of ecosystems	Ecosystems	Interpreting scientific evidence and data	C4	26
		Evaluating and designing scientific investigations	C4	27
		Explaining phenomena scientifically	C4	28
		Explaining phenomena scientifically	C4	29
		Evaluating and designing scientific investigations	C4	30

3.3. Development Stage

This stage is to produce a computer-based diagnostic assessment product to measure the student's readiness for junior high school to participate in AKM and PISA with automatic feedback that is valid and fit for use in schools. The development phase includes:

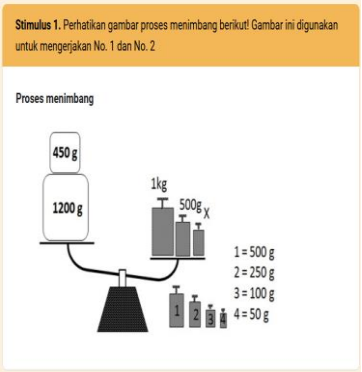
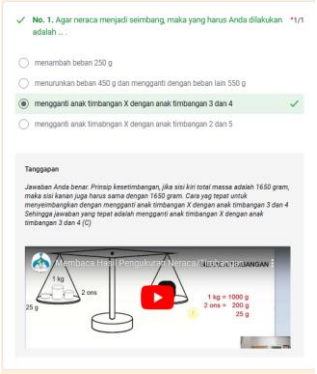
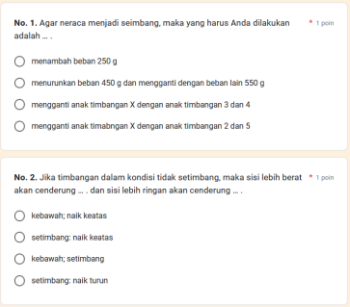
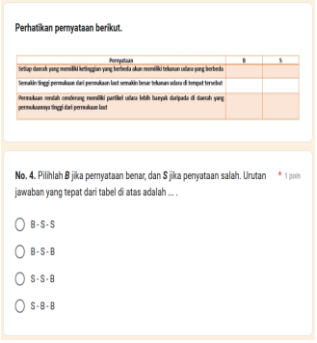
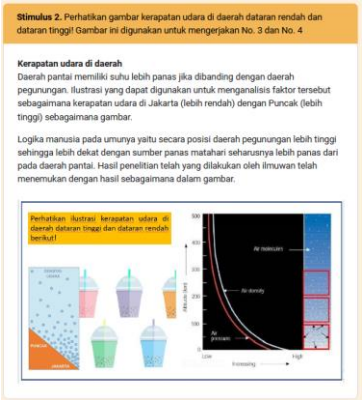
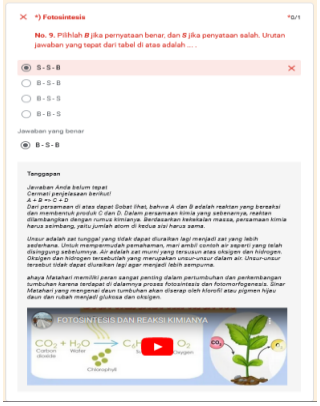
a. Making a diagnostic assessment question instrument with AKM and PISA models based on google form and completed by automatic feedback,

b. Product validation by experts, covering the aspects of construct, content, and language, and

c. Field test to students' junior high school at SMP N 1 Tempel and SMP N 4 Pakem Yogyakarta.

The results of stages (b) and (c) are used as the basis for revising the developed product. Table 6 presents the samples of the developed products.

Table 6. Product Display Example of Computer-Based Diagnostic Assessment.

Display	Description	Display	Description
	Stimulus Process of weighing		Automatic feedback model in the form of writing and video
	Multiple choice questions		Correct and Wrong statement model questions
	Stimulus images and graphics		Automatic feedback model in the form of writing and video

The obtained data were analyzed quantitatively and qualitatively. Details of data analysis in this study are explained as follows.

a. The validity of the research instrument describes that the product can describe and measure scientific literacy from the AKM and PISA model questions. Validity analysis includes constructs, content, and language, provided by expert judgments for each item of the product and analyzed using the Aiken formula [19]. The validity of the instrument is estimated through a feasibility test or content relevance, namely rational analysis by a competent panel or through an expert judgment consisting of constructs, content, and language. The results of the content validity were assessed by five science education experts with an assessment of five categories, namely very good (VG), good (G), fair (F), poor (P), very poor (VP) using the Aiken formula are described as in Table 7.

Table 7. The Analysis Results of the Validity of the AKM and PISA Model Questions with the Aiken Formula.

Assessment Aspects of the Expert	Expert's Assessment Results					$V_{calculate}$	Results
	1	2	3	4	5		
Content							
Questions following to the indicator	5	4	5	4	5	0,9	Valid
The answer choices are homogeneous and logical in terms of material	5	4	5	5	5	0,95	Valid
Questions have one correct or most correct answer.	5	5	4	4	5	0,9	Valid
Construct							
The main question is formulated clearly and unequivocally.	5	4	5	4	5	0,9	Valid
The main questions and answer choices are required statements only.	5	4	5	5	5	0,95	Valid
The main questions do not provide clues to the correct answer.	5	5	4	4	5	0,9	Valid
The main question does not contain statements that are double negative.	5	4	5	4	5	0,9	Valid
The length of the answer choice formulations is relatively the same.	5	5	5	5	5	1	Valid
The answer choices do not contain the statement, "All of the answer choices are wrong", or "All of the answer choices are correct".	5	5	4	5	5	0,95	Valid
The answer choices in the form of numbers or time are arranged order based on the value of the number or chronologically.	5	4	5	4	5	0,9	Valid
Pictures, graphs, tables, diagrams, and the like contained in the questions are clear and functional.	5	4	5	5	5	0,95	Valid
The items do not depend on the answers of the previous questions.	5	5	4	4	5	0,9	Valid
Language							
The question uses language following the rules of the Indonesian language.	5	4	5	4	5	0,9	Valid
The feedback does not use the local language.	5	5	5	5	5	1	Valid
The feedback uses communicative language.	5	5	4	4	5	0,9	Valid
The answer choices do not repeat words or phrases that are not a unified meaning.	5	4	4	4	5	0,85	Valid

Determination of valid or invalid based on Aiken table for 5% or p value < 0,05 is 0,80.

Table 8. The Results of Validity Analysis of the AKM and PISA Feedback Model with the Aiken Formula.

Assessment Aspects of the Expert	Expert's Assessment Results					$V_{calculate}$	Results
	1	2	3	4	5		
Content							
Feedback following to the indicator	5	4	5	4	5	0,9	Valid
Logical feedback in terms of material developed into questions	5	5	4	5	5	0,95	Valid
Construct							
Feedback is formulated clearly and unequivocally.	5	4	5	4	5	0,9	Valid
Feedback is a required statement only.	5	5	4	5	5	0,95	Valid
Explanation via video as feedback amplifier works	5	5	4	4	5	0,9	Valid
Language							
The question uses language following the rules of the Indonesian language.	5	4	5	4	5	0,9	Valid
The feedback does not use the local language.	5	5	5	4	5	0,95	Valid
The feedback uses communicative language.	5	5	4	4	5	0,9	Valid

b. Scale conversion analysis of the quantitative to qualitative results of experts related to the feasibility of the product from the content, construct, and language aspects with the provisions of the lowest score criteria put score as one and the highest score of five. The assessment of items based on the feasibility instrument shows that the mean ideal score of the three aspects of the device (Xi) = ½ (ideal maximum score + ideal minimum score) = 48.00. Standard Deviation of the ideal score of the three aspects (SBi) = ⅙ (ideal maximum score-ideal minimum score) = 10,67 is presented in Table 9.

Table 9. Score Conversion Scale to Question Item Criteria.

Interval Score	Score	Criteria
$X > \bar{X}_i + 1{,}80SB_i$	X > 67,21	Very Good (VG)
$\bar{X}_i + 0{,}60SB_i < X < \bar{X}_i + 1{,}80SB_i$	54,40 < X < 67,21	Good (G)
$\bar{X}_i - 0{,}60SB_i < X < \bar{X}_i + 0{,}60SB_i$	48,00 < X < 54,40	Fair (F)
$\bar{X}_i - 180SB_i < X < \bar{X}_i - 0{,}60SB_i$	28,79 < X < 48,00	Poor (P)
$X < \bar{X}_i - 1{,}80SB_i$	X < 28,79	Very Poor (VP)

Expert's assessment after conversion obtained the results in table 10.

Table 10. The Result of Score Conversion Scale to Question Item Criteria.

Development Product	Total Score (X)	Result	Decription
Items of Questions	75,00	X > 67,21	Very Good

Feedback assessment based on the instrument fit shows that the mean ideal score of the three aspects of the device (Xi) = ½ (ideal maximum score + ideal minimum score) = 24,00. Standard deviation of the ideal score of the three aspects (SBi) = ⅙ (ideal maximum score-ideal minimum score) = 5,33, presented in Table 10.

Table 11. Score Conversion Scale to Feedback Criteria.

Interval Score	Score	Criteria
$X > \bar{X}_i + 1{,}80SB_i$	X > 33,59	Very Good
$\bar{X}_i + 0{,}60SB_i < X < \bar{X}_i + 1{,}80SB_i$	27,20 < X < 33,59	Good
$\bar{X}_i - 0{,}60SB_i < X < \bar{X}_i + 0{,}60SB_i$	24,00 < X < 27,20	Average
$\bar{X}_i - 180SB_i < X < \bar{X}_i - 0{,}60SB_i$	14,41 < X < 24,00	Poor
$X < \bar{X}_i - 1{,}80SB_i$	X < 14,41	Very Poor

Expert assessment after conversion obtained the following results, presented in table 12.

Table 12. The results of Score Conversion Scale to Feedback Criteria.

Development Product	Total Score (X)	Result	Decription
Feedback	37,40	X > 33,59	Very Good

c. Item fit analysis with the Rasch model (1 PL model) using the QUEST application. The fit of items with the model in the QUEST program is based on the mean value of INFIT Mean of Square (INFIT MNSQ) and its standard deviation or the mean value of INFIT Mean of Square (INFIT MNSQ) or INFIT t. The fit of each item with the model in the QUEST program is based on the INFIT MNSQ value or the INFIT t item value presented in Table 2. The results of the analysis using the QUEST application from a limited trial of 75 students in grades VII, VIII, and IX at SMP N 1 Tempel are presented in Table 13.

Table 13. The Results of Item Fit Analysis with the QUEST Application.

Items	INFT MNSQ	OUTFT MNSQ	Description
1	1,05	1,43	Fit Model Rasch
2	1,03	1,28	Fit Model Rasch
3	0,99	0,98	Fit Model Rasch
4	1,09	1,09	Fit Model Rasch
5	1,2	1,28	Fit Model Rasch
6	1,09	1,19	Fit Model Rasch
7	1	1,07	Fit Model Rasch
8	0,97	1,07	Fit Model Rasch
9	1,18	1,17	Fit Model Rasch
10	0,91	0,89	Fit Model Rasch
11	1	0,99	Fit Model Rasch
12	0,97	0,9	Fit Model Rasch
13	1	0,98	Fit Model Rasch
14	1,07	1,2	Fit Model Rasch
15	0,77	0,74	Fit Model Rasch
16	0,93	0,83	Fit Model Rasch
17	1	0,94	Fit Model Rasch
18	0,79	0,75	Fit Model Rasch
19	0,88	0,73	Fit Model Rasch
20	0,95	0,86	Fit Model Rasch
21	1	1,09	Fit Model Rasch
22	0,94	1,09	Fit Model Rasch
23	0,9	0,81	Fit Model Rasch
24	1,06	1,21	Fit Model Rasch
25	0,99	2,17	Fit Model Rasch
26	1,01	1,1	Fit Model Rasch
27	1,04	2,27	Fit Model Rasch
28	1	1,13	Fit Model Rasch
29	0,87	0,87	Fit Model Rasch
30	1,02	1,61	Fit Model Rasch
Mean	0,99	1,12	Fit Model Rasch
SD	0,09	0,36	

d. Descriptive statistical analysis using the SPSS version 25 for Windows application to see the results of the trial to measure the student's readiness of class VIII in SMP N 1 Tempel and SMP N 4 Pakem to participate in AKM and PISA, presented in Table 14.

Table 14. Descriptive Statistical Analysis Results.

SMP N 4 Pakem		SMP N 1 Tempel	
Mean	20,38	Mean	14,63
Median	21,00	Median	15,00
Mode	22,00	Mode	15,00
Standard Deviation	3,73	Standard Deviation	5,06
Range	20,00	Range	22,00
Minimum	7,00	Minimum	7,00
Maximum	27,00	Maximum	29,00
Sum	652,00	Sum	468,00
Count	32,00	Count	32,00

Descriptive statistical analysis for the average scores of the two schools is presented in the Figure 2.

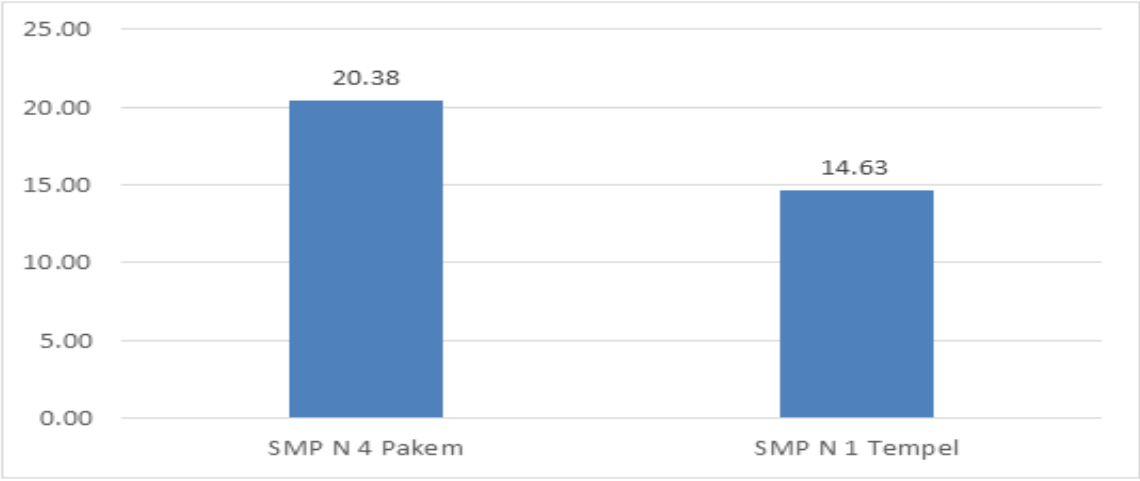


Figure 2. Graph of the Mean Score of Student's Readiness of Class VIII to Participate in AKM and PISA.

e. Analysis of student's readiness from the results of the trial was carried out with the percentage, presented in table 15.

Table 15. Results of Readiness Analysis to Participate in AKM and PISA.

School	Mean	Percentage	Description
SMP N 4 Pakem	20,38	67,92	Average
SMP N 1 Tempel	14,63	48,75	Very poor

3.4. Disseminate Stage

This stage is the stage of introducing the developed product on a wider scale which is presented in regional-level science teacher workshops and international seminars. Dissemination through international science education seminars attended by various science education circles, such as teachers, lecturers, and science education students. Wider distribution of the product was carried out in a science teacher workshop at SMP N 4 Pakem Yogyakarta offline, which was attended by 41 science teachers, and online using a zoom meeting as Figure 3.

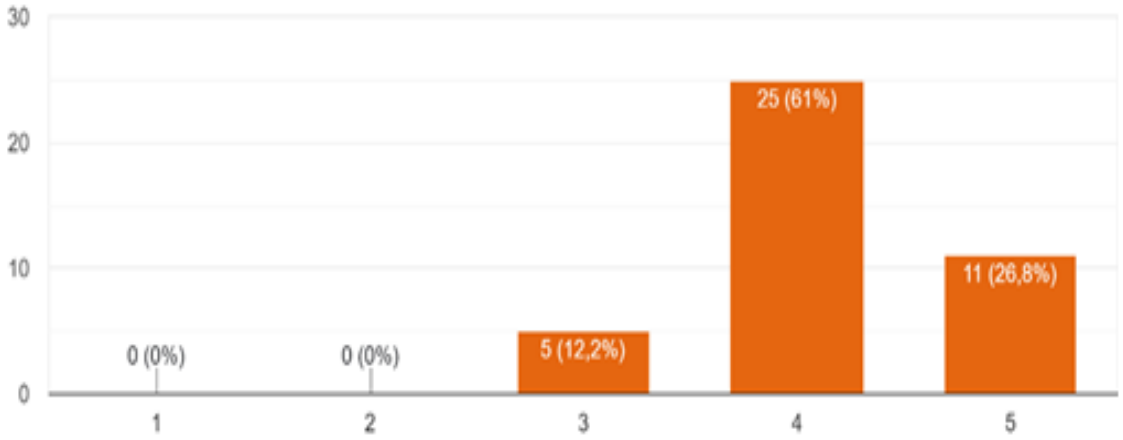


Figure 3. Assessment of Training Materials by Participants from the Aspect of Ease of Understanding.

Number description from the graph namely 5 is very good (VG), 4 is good (G), 3 is fair (F), 4 is poor (P), and 5 is very poor (VP). The results of the workshop as following.

a. Teachers have a basic understanding of the AKM and PISA model assessments and their application in the evaluation of science learning.

b. The teacher can arrange the AKM and PISA models of science questions.

In summary, the results of the teacher's responses might be concluded as good and support the competence of teachers to develop PISA model science items.

4. CONCLUSION

The conclusion of this study showed that research and development with a 4-D model are effectively used in developing computer-based diagnostic assessment products to measure the student's readiness in junior high school to participate in AKM and PISA with automatic feedback. In detail, it is summarized as follows.

1. A computer-based diagnostic assessment application was formed to measure the student's readiness of junior high school to participate in AKM and PISA with valid automatic feedback based on the Aiken formula and fit according to science education experts, seen from the aspects of content, construct, and language.

2. The results of the empirical test from the application of computer-based diagnostic assessment to measure the student's readiness of junior high school to participate in AKM and PISA with automatic feedback stating that all items fit the Rasch model.

3. The results of the limited-scale test for the student's readiness to participate in AKM and PISA show that SMP N 4 Pakem has a readiness of 67.92% in the average category, while SMP N 1 Tempel has a readiness of 48.75% in the category of below average.

REFERENCES

[1] H. David, "The PISA Results in mathematics and science: A comparison between Israel and Turkey. Journal for the Education of Gifted Young Scientists," *J. Educ. Gift. Young Sci.*, vol. 3, no. 1, pp. 22–28, 2015, doi: http://dx.doi.org/10.17478/JEGYS.2015111087.

[2] P. Turiman, J. Omar, A. M. Daud, and K. Osman, "Fostering the 21st Century Skills through Scientific Literacy and Science Process Skills," *Procedia - Soc. Behav. Sci.*, vol. 59, pp. 110–116, Oct. 2012, doi: 10.1016/j.sbspro.2012.09.253.

[3] OECD, "Reviews of National Policies for Education in Indonesia," Paris, 2015.

[4] M. Erfan and T. Ratu, "Pencapaian HOTS (Higher Order Thinking Skills) Mahasiswa Program Studi Pendidikan Fisika FKIP Universitas Samawa," *J. Pendidik. Fis. Dan Teknol.*, vol. 4, no. 2, pp. 208–212, Nov. 2018, doi: 10.29303/jpft.v4i2.831.

[5] A. Asrijanty, *AKM dan Implikasinya pada Pembelajaran*. Jakarta, Indonesia: Kemdikbud, 2020.

[6] D.M. Andikayana, N. Dantes, and I.W. Kertih, "Pengembangan Instrumen Asesmen Kompetensi Minimum (AKM) Literasi Membaca Level 2 Untuk Siswa Kelas 4 SD," *J. Penelit. Dan Eval. Pendidik. Indones.*, vol. 11, no. 2, pp. 81–92, Oct. 2021, doi: 10.23887/jpepi.v11i2.622.

[7] A. Asrijanty, *AKM dan Implikasinya pada Pembelajaran*. Jakarta, Indonesia: Pusat Asesmen dan Pembelajaran, 2020.

[8] Ardiana, Nurazmi, and Ariana, "Pelatihan Pengembangan Digital Assessment Bagi Guru-Guru Di Mts Muhammadiyah Mandalle Kecamatan Bajeng Barat Kabupaten Gowa," *J. Pengabdi. Kpd. Masy. Membangun Negeri*, vol. 5, no. 2, pp. 183–192, 2021, doi: https://doi.org/10.35326/pkm.v5i2.1664.

[9] R. Lam, "Understanding Assessment as Learning in Writing Classrooms: The Case of Portfolio Assessment," p. 18, 2018.

[10] E. T. A. Sekyi, "Assessment, Student Learning and Classroom Practice: A Review," *J. Educ. Pract.*, vol. 7, no. 21, pp. 1–6, 2016, doi: https://doi.org/10.7176/JEP.

[11] D. Firmanzah, "Implementasi asesmen diagnostik dalam pembelajaran IPA Pada masa pandemi COVID-19 di SMP/MTs wilayah Menganti, Gresik," *PENSA E-J. Pendidik. Sains*, vol. 9, no. 2, pp. 165–171, 2021, [Online]. Available: https://ejournal.unesa.ac.id/index.php/pensa/article/view/37361

[12] R. Millar and V. Hames, "Using Diagnostic Assessment to Enhance Teaching and Learning A Study of the Impact of Research-informed Teaching Materials on Science Teachers' Practices," the UK Economic and Social Research Council (ESRC), London, English, 2003.

[13] Y. Sun and M. Suzuki, "Diagnostic Assessment for Improving Teaching Practice," *Int. J. Inf. Educ. Technol.*, vol. 3, no. 6, pp. 607–611, 2013, doi: 10.7763/IJIET.2013.V3.345.

[14] S. Min and L. He, "Developing individualized feedback for listening assessment: Combining standard setting and cognitive diagnostic assessment approaches," *Lang. Test.*, vol. 39, no. 1, pp. 90–116, Jan. 2022, doi: 10.1177/0265532221995475.

[15] D. C. Rohim, S. Rahmawati, and I. D. Ganestri, "Konsep Asesmen Kompetensi Minimum untuk Meningkatkan Kemampuan Literasi Numerasi Siswa Sekolah Dasar," *J. Varidita*, vol. 33, no. 1, pp. 52–62, 2021, doi: 10.23917/varidika.v33i1.14993.
[16] D. Setyawarno, D. Rosana, and E. Widodo, "Kompetensi Guru IPASMPdi Kecamatan Mlati dalam Mengerjakan Soal IPAModel PISA," *Pancasakti Sci. Educ. Journa*, vol. 3, no. 2, pp. 91–98, 2018, doi: doi.org/10.24905/psej.v3i2.105.
[17] P. Paidi, *Metodologi Penelitian Pendidikan Biologi*. Yogyakarta, Indonesia: UNY, 2011.
[18] B. Subali, *Metodologi Penelitian Pendidikan Biologi*. Yogyakarta, Indonesia: UNY Press, 2019.
[19] D. Setyawarno, *Analisis skor*. Yogyakarta: FMIPA Universitas Negeri Yogyakarta, 2020.
[20] E. . Sinambela, S. . Manik, R. . Siregar, L. W. . Manurung, and C. N. . Saragi, "An Analysis of Translation Strategies Used by Students in Translating Idioms at the Second Grade of SMA Kalam Kudus", International Journal of Membrane Science and Technology, vol. 10, no. 2, pp. 550-557, 2023..
[21] R. K. Hambleton, H. Swaminathan, and H. J. Rogers, *Fundamentals of item response theory*. Newbury Park, California: Sage Publication, 1991.
[22] B. Subali and P. Suyata, *Pengembangan item tes konvergen dan divergen dan penyelidikan validitasnya secara empiris*. Yogyakarta: Diandra, 2012.
[23] D. Rosana and D. Setyawarno, *Statistik terapan untuk penelitian pendidikan*. Yogyakarta, Indonesia: UNY Press, 2016.
[24] A. Bhattacherjee, *Social Science Research: Principles, Methods, and Practices*. Tampa, Florida, USA: University of South Florida, 2012.
[25] J. W. Creswell, *Research design : qualitative, quantitative, and mixed methods approaches*. California: SAGE Publications, 2014.
[26] F. Faul, E. Erdfelder, A.-G. Lang, and A. Buchner, "G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences," *Behav. Res. Methods*, vol. 39, no. 2, pp. 175–191, May 2007, doi: 10.3758/BF03193146.
[27] A. Kurniawati, "Science Process Skills and Its Implementation in the Process of Science Learning Evaluation in Schools.," *J. Sci. Educ. Res.*, vol. 5, no. 2, p. 5, 2021, doi: https://doi.org/10.21831/jser.v5i2.44269.
[28] H. Hestiana and D. Rosana, "The Effect of Problem Based Learning Based Sosio-Scientific Issues on Scientific Literacy and Problem-Solving Skills of Junior High School Students," *J. Sci. Educ. Res.*, vol. 4, no. 1, pp. 15–21, Sep. 2020, doi: 10.21831/jser.v4i1.34234.

DOI: https://doi.org/10.15379/ijmst.v10i2.1411