Pegem Journal of Education and Instruction, Vol. 11, No. 4, 2021 (pp. 1-8)

RESEARCH ARTICLE WWW.PEGEGOG.NET

Developing an Instrument to Assess Students' Problem-Solving Ability on Hybrid Learning Model Using Ethno-STEM Approach through Quest Program

Fine Reffiane[1*], Sudarmin[2], Wiyanto[3], Sigit Saptono[4]

[1]College Students a Postgraduate Program Universitas Negeri Semarang, Indonesia

[2-4]Lecturer, A Postgraduate Program Universitas Negeri Semarang, Indonesia

Abstract

The purpose of the study was to develop an instrument for the problem solving ability of students in Hybrid learning by Ethno-STEM via Four-D Model. In the study, the first three stages of the Four-D Model namely; defining, designing, and developing were applied. As the participants, 143 students of elementary education in the 4th semester took part in the study. The participants were selected randomly from 30 universities in Central Java, Indonesia. The data were collected via the assessment instrument of problem solving using Ethno-STEM test ability adapted from Plomp model (2010). The instrument was composed of 9 essay type questions. The analysis of validity was conducted by 10 instructors and teachers, and the instrument was assessed valid and reliable. Empirical validity showed 0.6 for Aiken index, classified into medium validity. Meanwhile, Quest gained ± 2.0 for t INFIT, classified into good instrument.

Keywords: Ethno-STEM, quest, hybrid-learning, problem solving skills

Introduction

Students in this 21st century are supposedly able to compete globally, decide carefully, accurately, logically, and systematically, and consider problems from different point of views (Gulec & Temel, 2015). According to (Kennedy *et al*., 2015), ability of problem solving will help science future teachers in learning process in this global era, therefore the current science learning paradigm is that educational program students should be taught to create and develop a valid and reliable instrument to measure the ability of problem solving and its implementation in learning process.

Pandemic era triggers students to be continually adaptive, whereas they are at home. To develop students' potential from home, thinking process which can be understood from any viewpoint is needed. Learning process using local discernment approach is required in order to make students keep loving their home and continually develop themselves. It is expected that technologies for distance education help students in that learning process. Learning Management System is one of the solutions in the process of distance education. Online learning during distance education help students in this pandemic era (Stevenson *et al*., 2015). The ability of solving problem should be one of the materials taught in the learning process. There are four factors influencing the performance of students' problem solving ability; knowledge, confidence, individual control, and social environment (Dostál, 2015). Low problem solving ability is caused by the lack of knowledge, motivation and emotional aspect, and the use of learning model in which all of those give impacts on students' problem solving ability (Çetin, 2020).

Hybrid learning has a high flexibility using Learning Management System which can reach out students' area. During Covid 19, learning type which is usable for students is required (Herde *et al*., 2016). Although each student is separated but they are still able to communicate virtually. Lecturers' ability of directing students to keep their ability in solving problems is needed. Therefore, learning model of Hybrid learning using local discernment and culture in each students' zone is needed.

Online learning began in March 2020. As for students independently, they must actively follow information updates on which platforms their courses will carry out online learning, assign assignments / quizzes, and also provide material. This learning technique fully adapts to the lecturers' policies for each course (Cotta Natale *et al*., 2020). Ethno-STEM is a local discernment-based learning approach which relates theories and skills in Science, technology, Engineering, and Mathematics. Ethno-STEM approach expectedly gives solutions in this pandemic era of Covid-19 (Aiken, W.,

Corresponding Author e-mail: finereffiane@upgris.ac.id,

https://orcid.org/0000-0001-9651-6164

How to cite this article: Reffiane F, Sudarmin, Wiyanto, Saptono S (2021). Developing an Instrument to Assess Students' Problem-Solving Ability on Hybrid Learning Model Using Ethno-STEM Approach through Quest Program. Pegem Journal of Education and Instruction, Vol. 11, No. 4, 2021, 1-8

Source of support: Nil

Conflict of interest: None.

DOI: 10.47750/pegegog.11.04.01

Received: 26.04.2021

Accepted: 03.06.2021 **Publication:** 01.10.2021

2015). Hybrid Learning Model using Ethno-STEM is one of learning models that makes students considerate local goods in their region. This learning model also can develop students' knowledge and improve their mathematic logical (Alam, 2021). This learning model development needs a supportive instrument. Appropriate instrument can help the process of study running well. Therefore, Quest is used as a technique to analyse items in promoting this instrument (Tabatabaee-Yazdi *et al*., 2018).

Education is the main need for every human being. Education functions to inherit, develop and build culture and civilization. Learning in Indonesia consists of formal and non-formal education levels. Formal education includes: elementary school, middle school, to college. Meanwhile, non-formal education includes: early childhood education, youth education, female empowerment education, skills education and job training. The educational process in learning activities or in the classroom will run smoothly, conducive, and interactive if education can be carried out properly when the curriculum becomes the main support in the learning process. The curriculum contains many constructive elements so that learning runs optimally. Thus, the curriculum plays an important role for the success of an education. Education in schools consists of subjects that follow the applicable curriculum. One of these subjects is Science (Sudarmin *et al*., 2020).

Related to the evaluation of Hybrid learning result by Ethno-STEM, an instrument is needed to assess students' problem solving ability as future elementary teachers. Therefore, this study is conducted to develop a valid and reliable instrument to assess problem solving ability of students who are studying using Hybrid learning by Ethno-STEM.

The research question of the study is: How can an instrument be developed to assess students' problem-solving ability on Hybrid Learning Model Using Ethno-STEM Approach through Quest Program ?

2. Research Method

2.1 Research design

The study focused on an instrument administered to analyse students' problem solving skills of hybrid learning via ethno-STEM in the Quest program. The study adopted the development model of Four-D by Thiagarajan, Semmel & Semmel (1974). In the study, only three stages of four stages were implemented in this development; defining, designing, and developing. Defining stage was done to determine crucial aspects in solving problems. Meanwhile, in designing stage, the activities are relating the crucial aspects in solving problems with the characteristics of Ethno-STEM, determining the question form of the instrument, designing instrument based on the selected form and designing the questions outline of the instrument. Further, doing expert appraisal and developmental testing was done in developing stage.

2.2 Data collection tools

All tools supporting a study is named a study instrument, instrument of data collection. Data collection instruments are those utilized by the researchers to collect the data in order to be systematic (Onajite, 2021). In other words, an instrument is used to ease gaining data for the researcher related to the research problems by combining cognitive aspects with questions to stimulate. to get data accuracy, researchers should take some attention 1) research problems should be clear and specific 2) data resources or information should be clear 3) the instrument should be objective and validate 4) the data type should be clear and instrument should be easy to apply.

The data were collected via the assessment instrument of problem solving using Ethno-STEM test ability adapted from Plomp model (2010). The test was composed of 9 essay type questions aiming to collect, analyse, and present the data systematically and objectively concerning the Hybrid learning by Ethno-STEM. The test was developed upon 10 experts' appraisal based on a validation sheet as the guidance. The validation sheet was used to assess material aspects, construction, and language. Empirically, the instrument was tested online to 143 students of elementary education on 4th semester.

2.3 Participants

The participants were composed of 143 (49 males, 94 females) randomly selected students at the Departments of Elementary Education on their 4th semester from 30 different universities in Central Java, Indonesia.

2.4 Validity & Reliability

The research stages were (1) initial research; (2) prototype construction, (3) assessment stages. Initially, an analysis of learning achievement on science subject was done. Then, constructing prototype was carried out by creating items and scoring guidance (Abubakar, 2021). The last stage was assessment. It was conducted through expert validation, revision of test instrument, and content and construct validity test. There were five validators in this research. The subject of the study was 40 elementary teachers-to-be. The technique of data analysis was validating the content using Aiken index (1980). Aiken' V Index has range 0-1 which has been agreed based on the expert assessment towards the item match with measured construct psychology (Melo, 2018). The following is the formula of calculating Aiken index;

$$V = \sum(ri - lo)/[n(c-1)] \quad .(1)$$

Notes:

r = assessor's score
lo = score for low validity
c = score for high validity
n = number of experts conducting assessment
i =
n = number of assessors

Construct validity of instrument used theory of PCM. PCM is a model of 1-PL with small sample, around 30-300 (Keeves & Masters (1999) in (Subali *et al*., 2021)) and the limit of INFIT t is -2 to +2, therefore an item is not fit based on RM if the value is <-2,0 or > +2,0 with the error limit 5% (Herde *et al*., 2016). This study used model of 1 PL RM and PCM from QUEST program. QUEST program provides information to determine the validity of the test items based on PCM theory (Subali *et al*., 2021). QUEST program provides point values of the test items. Guilfod (1956) cited by (Aiken, W., 2015) confirmed that the level of instrument reliability is 0,80 < r < 1,00, grouped into very high reliability; 0,60 < r < 0,80 is categorized into high reliability; 0,40 < r < 0,60 (fair reliability); 0,20 < r< 0,40 (low reliability); and -1,00 < r < 0,20 is classified into very low reliability (unreliable).

3. Procedure and Findings

Initially, what was done was doing literature review about assessment of problem solving ability in science learning and doing an achievement analysis of basic concept of science course for elementary students-to-be. From the review, it was figured out that problem solving is a cognitive skill used to solve problems related to real life (Reffiane *et al*., 2018) using new, creative, systematic, and analytic ways (Sumarni *et al*., 2016). Problem solving skill deals with a skill to think critically, think analytically, and create productively in which all of them involve ability of quantifying, communicating, and giving responds critically (Nerland, 2018). Polya (1945) cited by Selcuk *et al*. (2018) described that problem solving has four phases; (1) Understanding problems; students will not be able to solve problems if they do not know the problems well. (2) Planning solutions; this phase depends on students' experience. The more varieties they experience problems, the more creative they will be in solving the problems. (3) Solving problems; it is a phase where designed plans to solve problems are executed. (4) Crosschecking; Students make sure all phases are completely done (Shute *et al*., 2016).

Every educational expert has different ways to measure problem solving skill (Marshall & Harron, 2018) developed an instrument of multiple choice to measure problem solving ability. Meanwhile, (Corbi & Burgos, 2017)) applied multi complex systems. Multiple complex system relates dependent

Table 1. The relationship of the basic concept of science and Ethno-STEM

No	Review of Science Concept	a. Ethno-STEM of Javanese in Creating Brebes Terasi (shrimp paste), Smoking process of Kendal fish, and Juwana Milkfish Aquaculture
1	b. The unit of Mass, length, duration, and temperature c. Measurement of mass, length, duration, and temperature	a. Composition: brown sugar, acetes, and salt in which per 100 kg of acetes needs 1 kg of brown sugar and 20 kg of salt to make terasi. b. The process of creating a pond with a width of 2.5 meters and a height of 0.5 meters. c. The smoking process of stingray is at 70o- 100oC. d. The duration for fish smoking is 60 minutes using heat smoking method e. The duration for drying acetes is a half day (not standard) f. Giving compose in new fish pond is as much as one fist of adult (not standard) g. Milk fish seeds (nener in Javanese) are maintained by farmers to be"fingerling/ *gelondongan*" (not standard measurement).
2	a. Ingredients, compound, composition, chemical characters, physics characters, physics conversion, chemical conversion	a. Ingredients to make terasi are salt (NaCl), brown sugar contains sucrose ($C_6H_{22}O_{11}$). The disposition elements of those compounds can be spelled out b. The use of salt before stingray smoking c. Terasi is a composition. d. Chemical characteristic of terasi, salt, and brown sugar e. Chemical shift on acetes as creating process f. Chemical shift on stingray
3	a. Temperature and measurement b. Heat movement c. Influences of temperature and heat shift toward objects	a. The heat temperature for smoking stingray uses firewood/coconut shell b. Smoking process use embers in which the heat transfer occurs through conduction c. Process of drying acetes and smoked fish using sunlight is a process of heat transfer through radiation.
4	a. Concept of alkali and its characteristics b. Concept of salt and its characteristics	c. The use of salt on the fish processed through smoking and on the terasi making d. The use of alkali compose on every creation of fish pond
5	a. Abiotic and biotic b. Food cycle c. Food web d. Food pyramid e. Predation	a. "*Klekap*" is a natural food pullulated by farmers. Klekap is a mixture of alga and decomposed organism under fishpond. b. Phytoplankton is a natural food which grows naturally in the fishpond. c. Alga as natural food is grown naturally in the fishpond. d. Milkfish is a living thing selected by farmers as aquaculture animal to make economical profit. e. Brackish is a mixture of seawater and underground water used as media for milkfish living. f. Food cycle of milkfish and phytoplankton/*klekap*.

items to be complex problems. The test of problem solving ability is developed based on four indicators of problem solving ability; (1) joining skills and using imagination, (2) developing model, (3) investigating, and (4) analysing data and making conclusion. Problem solving assessments conducted by Jonassen (2011) were schema problems, analogy, casual, and argument. Problem solving skills in this study are thinking process of understanding, planning, releasing plans, and evaluating problem plans (Smith, 1985).

The initial study done was reviewing the curriculum of Elementary School Teacher study program in four universities in Central Java and doing a study of ethno-STEM used in learning the basic concept of Science (Safdar & Idrees, 2021). The objectives of curriculum review were to find out the fit problem among the concept of science, review of ethno-STEM, and learning achievement (Ogungbade Aderonke & Ipadeola Oluwaseun, 2021). The following table shows the study result of the relationship between science and ethno-STEM (See Table 1).

The analysis result of the relationship between basic concept of science and ethno-STEM becomes the base in creating instrument of problem solving in learning science for the context of ethno-STEM (Philosophy & Naveed, 2021). Instrument of essay which bring about a variety of answers used to measure skill of problem solving of educational program students.

The stages of creating test are creating items and guidance of scoring of the test result in Table 2.

(Subali *et al*., 2021) stated that essay test with varied and levelled answer keys can be scored into three categories. For example, testers are invited to tell two factors causing an incident. Accordingly, the testers' answers are categorized into three; category 1 for score 0 (tester does not answer), category 2 for score 1 (tester gives one correct causing factor), category 3 for score 2 (tester states one correct causing factor). Table 3 in this study, scoring guidance of problem solving skill test in science learning for ethno-STEM context is as follows.

Test Stages

The stages of the test were testing content validity and construct validity. Construct validity examines the instrument quality to get expertized validators agreement toward questions with the variables going to be measured. Score of validator agreement level is presented below in table 4.

The stage of construct validity was done after the instrument revision based on the result of content validity test. The instrument was contributed to 143 educational program students. The construct validity on the instrument which previously had been done by the testers was done using QUEST program based on IRT theory and PCM approach. The result of instrument analysis is as follows in table 5

Based the table presented, item 1-40 was about 0.77-1.20 which is consequently resulted that those items fit validity test of 1 PL model in table 6.

Based on the table above, INFIT MNSQ score of all respondents is between 0.77-1.22. Hence, items fit with testers in table 7.

From the table, it is revealed that the TAU score is +- in which the difficulty level is classified into medium.

Discussion

The number of testers who joined in the examination of instrument was 143 teacher-to-be. As considered, the analysis process was done using modern theory, IRT (Items Responses Theory). In the use of IRT, some do calibration based on one parameter which is only based on difficulty level (symbolized to β or b); therefore, called one logistic parameter or 1-PL model

Table 2: Instrument of Problem Solving Test

Dimension	*Indicators of Problem Solving Questions*	*Questions*
Understanding problems	Students are able to write concept of Science to solve contamination problems of milk fish pond Juwana, Rembang terasi creation, and Kendal stingray smoking.	1 a,2a,3a
	Students are able to write questions related to contamination of Juwana milk fish pond, Rembang terasi creation, and Kendal stingray smoking.	1b,2b,3b
	Students are able to write factors causing contamination problems of Juwana milk fish pond, Rembang terasi creation, and Kendal stingray smoking which should be finished.	1c,2c,3c
	Students are able to write essential information related to contamination problems of Juwana milk fish pond, Rembang terasi creation, and Kendal stingray smoking.	1d,2d,3d
Planning solutions	Students are able to write the relationship between appropriate concept of Science dealing with contamination problems of Juwana milk fish pond, Rembang terasi creation, and Kendal stingray smoking.	1e,2e,3e
Executing problem solving	Students are able to write concept of Science to solve contamination problems of milk fish pond Juwana, Rembang terasi creation, and Kendal stingray smoking.	1f,2f,3f
Crosschecking	Students are able to choose one solution about contamination problems of milk fish pond Juwana, Rembang terasi creation, and Kendal stingray smoking.	1g,2g,3g

Table 3: Scoring guidance of problem solving test

Number of questions	*21*
Maximum score	3x21=63
Minimal score	1x21=21
Criteria scale	63-21/3=14
Category of high problem solving skill	49 < X ≤ 63
Category of high problem solving skill	35 < X ≤ 49
Category of high problem solving skill	21 < X ≤ 35

Table 4: Data of Content Validity

		Scoring Average	
No	*Validator*	*Language Clarity*	*Theory Agreement*
1	I	4.0	4.0
2	II	4.0	4.3
3	III	4.0	4.7
4	IV	4.0	4.4
5	V	5.0	4.4
Average		4.2	4.4
Aiken (V) Index		0.7	0.84
Validity		Valid	Very Valid

Table 5: Instrument fit based on model

```
-2.0                          |
------------------------------------------------------------------------------------------------------------
  Each X represents    1 students
============================================================================================================
tes KR
------------------------------------------------------------------------------------------------------------
Item Fit                                                                                    1/ 3/21 9:55
all on all (N = 143 L = 40 Probability Level= .50)
------------------------------------------------------------------------------------------------------------
INFIT
 MNSQ          .56       .63       .71       .83      1.00      1.20      1.40      1.60      1.80
-------------+---------+---------+---------+---------+---------+---------+---------+---------+---------
  1 item 1                                  .          | *          .
  3 item 3                                  .          |*           .
  5 item 5                                  .        *  |            .
  7 item 7                                  .          |  *         .
  9 item 9                                  .         *|            .
 11 item 11                                 .        * |            .
 13 item 13                                 .        * |            .
 15 item 15                                 .          | *          .
 17 item 17                                 .          |       *    .
 19 item 19                                 .         *|            .
 21 item 21                                 .        * |            .
 23 item 23                                 .         *|            .
 25 item 25                                 .         * |            .
 27 item 27                                 .         *|            .
 29 item 29                                 .        * |            .
 31 item 31                                 .         *|            .
 33 item 33                                 .          |*           .
 35 item 35                                 .          |*           .
 37 item 37                                 .        *  |            .
 39 item 39                                 .         *|            .
============================================================================================================
```

Table 6. Test of item fit based on tester

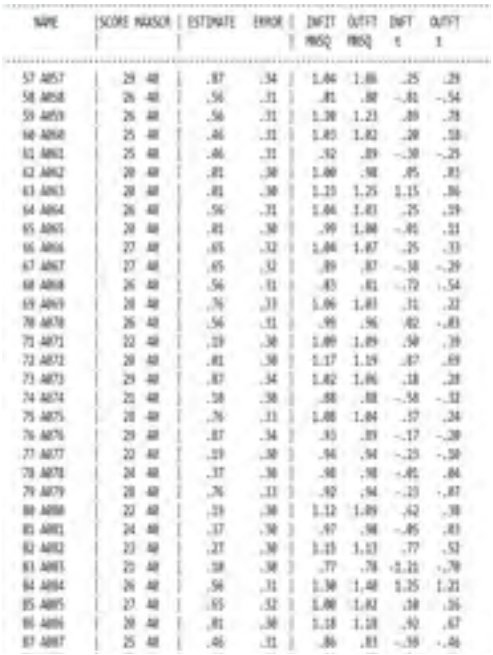


Table 7. Difficulty level of Item

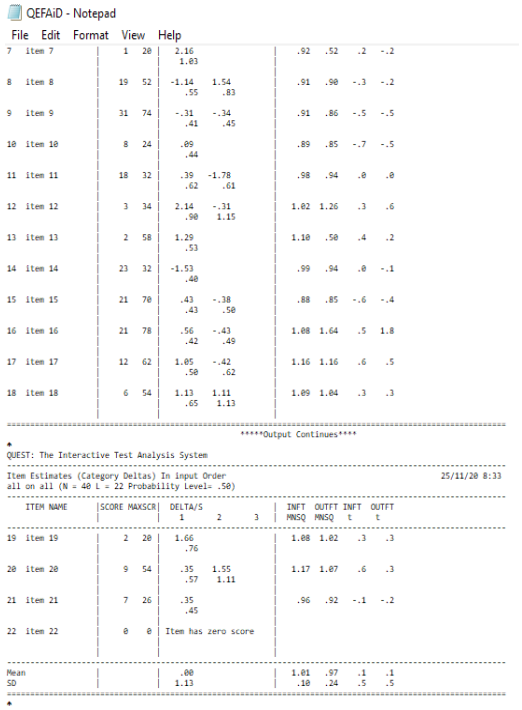
```
tes KR
------------------------------------------------------------------------------------------------------------
Item Estimates (Difficulty and Taus) In input Order                                         1/ 3/21 9:55
all on all (N = 143 L = 40 Probability Level= .50)
------------------------------------------------------------------------------------------------------------
    ITEM NAME       |SCORE MAXSCR| DIFFCLTY  TAU/S                          | INFT  OUTFT INFT  OUTFT
                    |            |            1      2      3      4      | MNSQ  MNSQ   t     t
------------------------------------------------------------------------------------------------------------
1   item 1          |   181  286 |   -.05    -.36    .36                   |  1.05  1.03   .6    .3
                    |            |    .12     .22    .17                   |
                    |            |                                         |
2   item 2          |     0    0 | Item has zero score                     |
                    |            |                                         |
                    |            |                                         |
3   item 3          |   188  286 |   -.18    -.50    .50                   |  1.02  1.00   .2    .0
                    |            |    .12     .24    .17                   |
                    |            |                                         |
4   item 4          |     0    0 | Item has zero score                     |
                    |            |                                         |
                    |            |                                         |
5   item 5          |   171  286 |    .06    -.48    .48                   |   .93   .93  -.8   -.6
                    |            |    .12     .22    .17                   |
                    |            |                                         |
6   item 6          |     0    0 | Item has zero score                     |
                    |            |                                         |
                    |            |                                         |
7   item 7          |   157  286 |    .24    -.49    .49                   |  1.05  1.04   .6    .4
                    |            |    .12     .20    .18                   |
                    |            |                                         |
8   item 8          |     0    0 | Item has zero score                     |
                    |            |                                         |
                    |            |                                         |
9   item 9          |   179  286 |    .01    -.16    .16                   |   .99   .98  -.2   -.2
                    |            |    .11     .21    .17                   |
                    |            |                                         |
10  item 10         |     0    0 | Item has zero score                     |
```

also Rasch Model. 1-PL model, in a research, can be used for small sample (Shute *et al*., 2016). Some experts said that the specific sample number for 1-PL model of Rasch Model is 30 to 300 with the limit of INFINIT t -2 to +2 (Nerland, 2018). So that, the error limit used is 5% to get ±1,96 for INFINIT t or rounded to ±2,0. Thereupon, according to Rasch Model an item is not fit when the value is <-2,0 or > +2,0 (probability <0,05). Before a test is used to know students' progress, the test should be prepared well. the characteristics of it should have been known, prepared, and designed based on the guidelines. after the test is applied, teachers can obtain feedback about the quality of the test by checking students' score.

The instrument used for the study was in the form of essay. The result of the essay examination can be analysed based on PCM. Some consideration of using PCM as extension of RM (1-PL model) is the use of sample is not as big as calibration of polytomous data using 2-PL or 3-PL model (Sudria *et al*., 2018). In addition, the response characteristics toward each item follow PCM in which the difficulty level of each category from the lower to the higher one is different, thus the value of delta for the lower category and the higher one is different as well (Cotta Natale *et al*., 2020).

Program used to do an instrument analysis was QUEST. QUEST provides information to determine the validity of an item based on PCM theory which is an extension of RM. The following is the steps of reading the analysis result (Librarian *et al*., 2021).

Table 8. Validity test to know the fit between the items and the tester toward the model



QEFAiD - Notepad

File Edit Format View Help

7	item 7	1	20	2.16 1.03		.92	.52	.2	-.2
8	item 8	19	52	-1.14 .55	1.54 .83	.91	.90	-.3	-.2
9	item 9	31	74	-.31 .41	-.34 .45	.91	.86	-.5	-.5
10	item 10	8	24	.09 .44		.89	.85	-.7	-.5
11	item 11	18	32	.39 .62	-1.78 .61	.98	.94	.0	.0
12	item 12	3	34	2.14 .90	-.31 1.15	1.02	1.26	.3	.6
13	item 13	2	58	1.29 .53		1.10	.50	.4	.2
14	item 14	23	32	-1.53 .40		.99	.94	.0	-.1
15	item 15	21	70	.43 .43	-.38 .50	.88	.85	-.6	-.4
16	item 16	21	78	.56 .42	-.43 .49	1.08	1.64	.5	1.8
17	item 17	12	62	1.05 .50	-.42 .62	1.16	1.16	.6	.5
18	item 18	6	54	1.13 .65	1.11 1.13	1.09	1.04	.3	.3

*****Output Continues****

QUEST: The Interactive Test Analysis System

Item Estimates (Category Deltas) In input Order 25/11/20 8:33
all on all (N = 40 L = 22 Probability Level= .50)

	ITEM NAME	SCORE	MAXSCR	DELTA/S 1	2	3	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
19	item 19	2	20	1.66 .76			1.08	1.02	.3	.3
20	item 20	9	54	.35 .57	1.55 1.11		1.17	1.07	.6	.3
21	item 21	7	26	.35 .45			.96	.92	-.1	-.2
22	item 22	0	0	Item has zero score						
Mean				.00			1.01	.97	.1	.1
SD				1.13			.10	.24	.5	.5

Item Characteristic Curve (ICC) will go horizontal if the value of INFINIT MNSQ for item or e is bigger than logit unit > 1,30 or <0,77. As a result, it will make platocurtic curve not leptokurtic curve anymore (Corbi & Burgos, 2017). Thus, QUEST program determines that an item or tester/case/person is considered fit with the model using limit of INFIT MNSQ from 0.77 to 1.30 (Azhari & Fajri, 2021). Somehow, some studies use more strict limitation which is about 0.83 to 1.20 while others do examination based on value of INFINIT t which use ± 2,0 (rounded to ± 1,96) for t value if the error level or alpha is 5% (Davis *et al*., 2019). According to Rasch Model, an item does not fit if its value is <-2,0 or > +2,0 (probability <0,05).

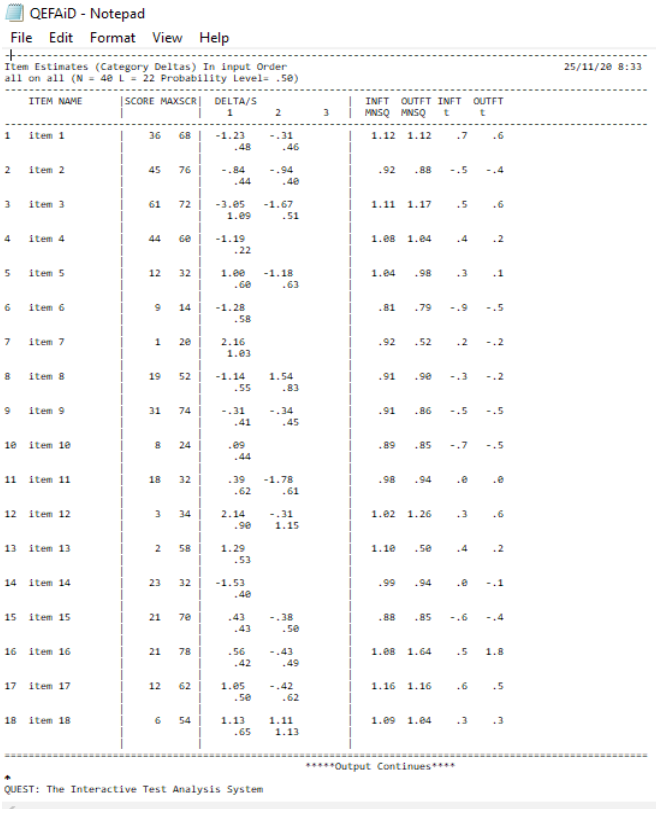
- Analysis of problem solving

The output result of problem solving test using QUEST program is presented below.

Based on the value of INFIT MNSQ, item number 1-6 are accepted since their INFIT MNSQ value is from 0.77 to 1.30 (Azhari & Fajri, 2021), therefore it fits based on model. Additionally, using INFIT t with the limit ±2,0 item 1-6 fits based on testers' ability. In this case, the t used is ± 2,0 (round of ± 1,96) if the error level or alpha is 5% (Herde *et al*., 2016). In accordance with that, item 1-6 for creative thinking test fits based on Rasch Model if the value is <-2,0 or > +2,0 (probability <0,05).

Based on the difficulty level and discrimination power, item 1-6 of creative thinking test is classified into difficult category. In this study, what is measured is the creative thinking skill, so that the purpose of the test is selective. Accordingly, items classified into very difficult can be used (Subali *et al*., 2021).

Table 9. Output of problem solving test



QEFAiD - Notepad

File Edit Format View Help

Item Estimates (Category Deltas) In input Order 25/11/20 8:33
all on all (N = 40 L = 22 Probability Level= .50)

	ITEM NAME	SCORE	MAXSCR	DELTA/S 1	2	3	INFT MNSQ	OUTFT MNSQ	INFT t	OUTFT t
1	item 1	36	68	-1.23 .48	-.31 .46		1.12	1.12	.7	.6
2	item 2	45	76	-.84 .44	-.94 .40		.92	.88	-.5	-.4
3	item 3	61	72	-3.05 1.09	-1.67 .51		1.11	1.17	.5	.6
4	item 4	44	60	-1.19 .22			1.08	1.04	.4	.2
5	item 5	12	32	1.00 .60	-1.18 .63		1.04	.98	.3	.1
6	item 6	9	14	-1.28 .58			.81	.79	-.9	-.5
7	item 7	1	20	2.16 1.03			.92	.52	.2	-.2
8	item 8	19	52	-1.14 .55	1.54 .83		.91	.90	-.3	-.2
9	item 9	31	74	-.31 .41	-.34 .45		.91	.86	-.5	-.5
10	item 10	8	24	.09 .44			.89	.85	-.7	-.5
11	item 11	18	32	.39 .62	-1.78 .61		.98	.94	.0	.0
12	item 12	3	34	2.14 .90	-.31 1.15		1.02	1.26	.3	.6
13	item 13	2	58	1.29 .53			1.10	.50	.4	.2
14	item 14	23	32	-1.53 .40			.99	.94	.0	-.1
15	item 15	21	70	.43 .43	-.38 .50		.88	.85	-.6	-.4
16	item 16	21	78	.56 .42	-.43 .49		1.08	1.64	.5	1.8
17	item 17	12	62	1.05 .50	-.42 .62		1.16	1.16	.6	.5
18	item 18	6	54	1.13 .65	1.11 1.13		1.09	1.04	.3	.3

*****Output Continues****

QUEST: The Interactive Test Analysis System

- Estimation of Reliability

Estimation of reliability, based on IRT, calculated based on items is called item separation index while calculated based on case or person is called person separate index. The higher the estimation of item separate index is, the more valid the analysis of all items is based used model (Whether based on RM, PCM, or RSM).

Instruments or data collection tools are tools used to collect data in a study. Research instruments are all tools used to obtain, manage, and interpret information from respondents which is carried out with the same measurement pattern. Research instruments are designed for one purpose and cannot be used in other studies (Kim, 2021).

The uniqueness of each research object causes a researcher to design his own instrument. The arrangement of instruments for each study is not always the same as for other researchers. This is because the goals and mechanisms of action in each research technique are different. The data collected using certain instruments will be described and attached or used to test the hypothesis proposed in a study. To collect data in a study, we can use existing instruments and can also use instruments that are made by ourselves. The available instruments are generally considered standard instruments to collect data on certain variables.

Table 10. The Good Instrument of Reliability

KRtn - Notepad

File Edit Format View Help

```
Item Analysis Results for Observed Responses                                    1/ 3/21 9:55
all on all (N = 143 L = 40 Probability Level= .50)
-------------------------------------------------------------------------------------------
.....................................................................................

Item   40: item 40                          Infit MNSQ =  .00
                                                  Disc =  .00

Categories          1        2        3        4        5        9      missing

Count                0        0        0        0        0        0      143
Percent (%)         .0       .0       .0       .0       .0       .0
Pt-Biserial         NA       NA       NA       NA       NA       NA
p-value             NA       NA       NA       NA       NA       NA
Mean Ability        NA       NA       NA       NA       NA       NA      .43

Step Labels

Thresholds
Error
.....................................................................................

Mean test score       24.55
Standard deviation     3.37
Internal Consistency    .04

The individual item statistics are calculated
using all available data.

The overall mean, standard deviation and internal
consistency indices assume that missing responses
are incorrect.  They should only be considered useful when
there is a limited amount of missing data.
=====================================================================================
```

This, if a standard instrument is available to collect research variable data, we can immediately use the instrument, provided that the theory on which the instrument is based is in accordance with the theory referred to in our research. In addition, the variable constructs measured by the instrument are the same as the variable constructs that we want to measure in our research. However, if a standard instrument is not yet available to collect data on these variables, the researcher must make it himself.

Research instruments are needed by researchers in collecting data when conducting research. This research instrument can make it easier for researchers to manage, analyze and draw conclusions from the data that has been collected. the research instrument is a measuring tool used to record the state and activity of psychological attributes. The data generated are generally in the form of quantitative data which technically can be classified into cognitive attributes and non-cognitive attributes. The cognitive attribute of the stimulant is a question, while the non-cognitive attribute of the stimulus is a statement. Data collection can be done in various places, various sources, and various ways. When viewed from the point of view of the place, research data can be collected in the laboratory using experimental methods, or at home with various respondents, and so on. When viewed from the data source, research data can be collected through primary and secondary data sources.

The higher the person separation index is the more consistent the assessor of each item used to measure related testers. Testers-based reliability estimation is the same as CTT-based reliability estimation – reliability using alpha Cronbach for polytomous data and using Kuder-Richradson-20 for dichotomous data. (Nerland, 2018) called item separation index as sample reliability, while person separation index as test reliability. QUEST program also provides the result of rest reliability based on CTT, in the form of internal consistency index in which for polytomous scoring use alpha Cronbach index while for dichotomous use KR-20 index(Davis *et al*., 2019). In this case, the valid reliability is test functioned as selection not achievement measurement.

Accurate means that the measurement can provide an overview of the smallest differences regarding the differences between one another. For example, in the field of physical aspect measurement, if we Want to know the weight of a gold ring, we must use a gold weighing device so that the weighing results are valid, that is, precise and accurate. A body weigher does measure weight, but it is not accurate enough to weigh a gold ring because a very small difference in weight in the weight of gold will not be seen on a weight gauge (Cotta Natale *et al*., 2020).

The reliability is 0.44. According to Guilfood (1956) cited by (Subali *et al*., 2021), this value based on reliability level is categorized into good, thus the reliability of the instrument is good and can be used for measurement. If the reliability is -1 to -2, it is unreliable.

- Conclusion and Recommendations

Based on the Aiken index, the language clarity of the items is 0.7 and categorized into valid, while the theory match gains 0.84 and considered good validity. The value of INFINIT t for 21 items is between ±2.0, therefore the instrument fits the items and the testers completes 1-PL model examination. The reliability of the instrument is 0.45 and considered good to be used to measure problem solving ability.

Quest, which has been used widely in education environment in Indonesia a computer program, is used to help in analysing questions using modern approach one parameter with its advantages and disadvantages. The materials taught in this study is how to operate QUEST and analyse the output of the program. The objective of the study is teachers have one of pedagogical competences; which is a competence of using technology in learning process. They should be able to operate QUEST for analysing questions. Quest needs longer time in the data processing, so that it takes longer time to discuss. Users should have an ability to read and analyse every item of the test measured. For classical test instrument, the ability covers difficulty level, discrimination power and distractor effectivity. meanwhile, modern test parameter is difficulty level of Rasch model, error, and INFIT MNSQ. The success of study instrument for Hybrid learning using Ethno STEM is expectedly expanding scholarship on education field.

References

Abubakar, H. O. (2021). Availability and Use of Audio-Visual Materials For Teaching Mathematics At The senior Secondary School in Ibadan, South-West Local Government, Oya State. *Library Philosophy and Practice*, 68–70.

Aiken, W., & J. (2015). A Preliminary Report of Advisor Perceptions of Advising and of a Profession. *Nacada Journal*, *35*(2), 60–70.

Alam, N. A. R. (2021). The Importance of Parent'S Literacy Understanding Towards Children Reading Habits. *Library Philosophy and Practice*, 1–10.

Azhari, B., & Fajri, I. (2021). Distance learning during the COVID-19 pandemic: School closure in Indonesia. *International Journal of Mathematical Education in Science and Technology*, 20–29.

Corbi, A., & Burgos, D. (2017). Open distribution of virtual containers as a key framework for open educational resources and STEAM subjects. *Electronic Journal of E-Learning*, *15*(2), 126–136.

Cotta Natale, C., Seixas Mello, P., Frateschi Trivelato, S. L., Marzin-Janvier, P., & Manzoni-de-Almeida, D. (2020). Evidence of Scientific Literacy Through Hybrid and Online Biology Inquiry-Based Learning Activities (Journal Pre-Proof). *Higher Learning Research Communications*, *11*, 33–49.

Çetin, A. (2020). Examining project-based STEM training in a primary school. International Online Journal of Education and Teaching (IOJET), 7(3). 811- 825.

Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., McMorran, R., Wiegers, J., Wiegers, T. C., & Mattingly, C. J. (2019). The Comparative Toxicogenomics Database: Update 2019. *Nucleic Acids Research*, *47*(D1), D948–D954.

Dostál, J. (2015). Theory of Problem Solving. *Procedia - Social and Behavioral Sciences*, *174*, 2798–2805.

Gulec, S., & Temel, H. (2015). Body Language Using Skills of Teacher Candidates from Departments of Mathematics Education and Social Studies Education. *Procedia - Social and Behavioral Sciences*, *186*, 161–168. https://doi.org/10.1016/j.sbspro.2015.04.149

Herde, C. N., Wüstenberg, S., & Greiff, S. (2016). Assessment of Complex Problem Solving: What We Know and What We Don't Know. *Applied Measurement in Education*, *29*(4), 265–277.

Kennedy, G., De Barba, P., Coffrin, C., & Corrin, L. (2015). Predicting success: How learners' prior knowledge, skills and activities predict MOOC performance. *ACM International Conference Proceeding Series*, *16-20-Marc*, 136–140.

Kim, Y. (2021). The Problem / Project-Based Learning (PBL / PjBL) at Online Classes. *International Journal Advanced Culture Technology*, *9*(1), 162–167.

Librarian, A., Asare, C., Librarian, A., Anafo, P., & Librarian, A. D. (2021). Information Literacy and Academic Libraries : Students learning in the information. *Library Philosophy and Practice*, *233*, 71–81.

Marshall, J. A., & Harron, J. R. (2018). Making Learners : A Framework for Evaluating Making in STEM Education The Interdisciplinary Journal of Problem-based Learning Special Issue : Tinkering in Technology-Rich Design Contexts Making Learners : A Framework for Evaluating Making in STEM Educatio. *Interdisciplinary Journal of Problem-Based Learning*, *12*(2).

Melo, M. (2018). The 4C/ID-model in physics education: Instructional design of a digital learning environment to teach electrical circuits. *International Journal of Instruction*, *11*(1), 103–122.

Nerland, M. (2018). Knowledge practices and relations in professional education. *Studies in Continuing Education*, *40*(3), 242–256.

Ogungbade Aderonke, A., & Ipadeola Oluwaseun, L. (2021). From Classroom to the Field of Library Practice: the Gaps and its Cost in the Librarianship Profession. *Library Philosophy and Practice*, *2021*, 1–15.

Onajite, R. (2021). Influence of Information Behaviour on Substance Use of Undergraduates in Universities in Ogun State, Nigeria. *Library Philosophy and Practice*, 1–16.

Philosophy, L., & Naveed, M. A. (2021). DigitalCommons @ University of Nebraska - Lincoln Evidence on Psychometric Properties of Scales Assessing Information Related Anxieties : A Systematic Review Evidence on Psychometric Properties of Scales Assessing Information Related Anxieties : A Systema. *Library Philosophy and Practice*, 55–64.

Reffiane, F., Sudarmin, M., Wiyanto, M., & Prasetyo, A. B. (2018). Blended Learning Analysis in Nature Basic Concept in Elementary School Teacher Education Study Program. *ISET*, *247*, 169–171.

Safdar, M., & Idrees, H. (2021). Assessing Undergraduate and Post Graduate Students' Information Literacy Skills: Scenario and Requirements in Pakistan. *Library Philosophy and Practice*, 50–61.

Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, *63*, 106–117.

Smith, W. J. (1985). Thinking Skills. In *NASSP Bulletin* (Vol. 69, Issue 480).

Stevenson, M., Hedberg, J., Highfield, K., & Diao, M. (2015). Visualizing solutions: Apps as cognitive stepping-stones in the learning process. *Electronic Journal of E-Learning*, *13*(5), 366–379.

Subali, B., Yogyakarta, U. N., Surakarta, U. M., Aminah, N. S., & Maret, U. S. (2021). *Modern Test Theory*. *14*(1), 647–660.

Sudarmin, Sumarni, W., Mursiti, S., & Sumarti, S. S. (2020). Students' innovative and creative thinking skill profile in designing chemical batik after experiencing ethnoscience integrated science technology engineering mathematic integrated ethnoscience (ethno-stem) learnings. *Journal of Physics: Conference Series*, *1567*(2).

Sudria, I. B. N., Redhana, I. W., Kirna, I. M., & Aini, D. (2018). Effect of Kolb's learning styles under inductive guided-inquiry learning on learning outcomes. *International Journal of Instruction*, *11*(1), 89–102. https://doi.org/10.12973/iji.2018.1117a

Sumarni, W., Sudarmin, Wiyanto, & Supartono. (2016). The reconstruction of society indigenous science into scientific knowledge in the production process of palm sugar. *Journal of Turkish Science Education*, *13*(4), 281–292.

Tabatabaee-Yazdi, M., Motallebzadeh, K., Ashraf, H., & Baghaei, P. (2018). Development and validation of a teacher success questionnaire using the rasch model. *International Journal of Instruction*, *11*(2), 129–144.