Development of an Instrument to Measure Students' Sustainability Awareness in Physics Learning

Aufani Hazimah Saptaji [1,*], Didi Teguh Chandra[1], Ida Kaniawati [1]

1 Indonesian Education University, Bandung, Indonesia

* Email: itsmeaufaniihazs@gmail.com

ABSTRACT

By integrating ESD concepts and science, the aim is to increase student awareness and develop student patterns that support sustainable development. Therefore, equipment is needed to measure this posture. This research aims to develop a desire awareness tool for middle school students regarding temperature and heat. The research method used is research and development using the ADDIE model. The ADDIE model consists of five stages: analysis, design, development, implementation, and evaluation. The tool developed consists of 15 statements regarding ESD dimensions (environmental, social, economic) and desire awareness categories (behavior and attitude awareness, emotional awareness, desire practice awareness). Based on the Rasch modeling results, it was found that most of the items were valid, with item reliability values in the "excellent" category and Cronbach's alpha values in the "very good" category. The good condition of the respondents and items makes us confident that this instrument can be used as a data collection tool. Andrich's threshold on the part moves towards the positive. This means that the choices given are valid for the respondent and the unidimensionality is included in the "specific" category. This means that the instrument used to measure desire awareness is not influenced by other variables.

KEYWORDS

ESD; Instrument development; Sustainabilit awarness; Physics learning

1. INTRODUCTION

Global warming, the spread of deserts, the biodiversity crisis, disruption of the ozone layer and tropical rainforests, and water and air pollution are environmental problems whose sustainability is already threatened (Segara, 2015). Education for sustainable development has been discussed widely since 1992 with the aim of preparing future generations to do it in an even better way. To encourage students to think critically and have awareness of sustainable values (sustainability awareness) in dealing with environmental problems, the government is making efforts to integrate knowledge about the environment with everyday life. ESD aims to give people the opportunity to make decisions and do things that can improve their standard of living without sacrificing the sustainability of the Earth (Tristananda, 2018). Sustainability awareness emerged as a result. Sustainability awareness is sustainable awareness related to the environment around students which motivates them to maintain and respect the environment and other life around them (Nursadiah, 2018). One way to increase sustainability awareness is to use an environmental learning (ESD) approach by selecting materials that are adapted to the three pillars of ESD (Nursadiah, 2018). A tool is needed that can measure students' sustainability awareness. The focus of this research is to create an instrument that can measure students' sustainability awareness. It combines three dimensions of research, namely economic, societal, and environmental (Atmaca et al., 2019), and also incorporates students' level of awareness of sustainable development concepts and practices, attitudes, and moral values for sustainability research (Hassan et al., 2010). Most students continue to believe that temperature and heat are complicated subjects. Examples of problems related to temperature and heat include their relationship to the environment, their applications and influence on sustainable growth, and how students perceive them after learning them in everyday life. As a result, instrument development concentrated on temperature and heat issues.

2. METHODOLOGY

In this research, research development (R&D) methods, also known as research and development, are used to create certain products and test how effective they are (Sugiyono, 2017). R&D research is longitudinal, meaning it consists of several stages determined by the selection of the development model (Sugiyono, 2017). Research and development (R&D) is long-term, which means it consists of several stages determined by the selected development model (Sugiyono, 2015).

To create this instrument, we chose the ADDIE model because this model is suitable for developing targeted and effective tools. There are five stages in the ADDIE model: analysis, design, development, implementation, and evaluation. The ADDIE model development stage has the advantage of being more logical and complete than other research models (Mulyatiningsih, 2012).

The type of research and development (R&D) method used in this research , as explained in the figure above . Therefore , the research procedures is as follows:

1. Analysis

   Needs analysis, problem identification, and tasks analysis are part of the sustainability awareness instruments development process. At this stage, researchers collect as much data as possible for basic competency studies , core competencies, and journaling studies. This stages begins by analyzing basic

competencies and cores competencies related to temperature and heat for high school level, which finds ideas related to ESD aspects . Next , the researchers looked at the journaling used as a reference for developing the instruments . This research is based on research conducted by Hassan et al. (2010) and Atmaca et al. (2019).

2. Design

When the researchers begins to create a plan, or blueprints , the design phase begins. At this stage, the researchers formulates the problem, determines respondents according to research needs , and creates an environmental awareness instruments designed based on the results of the researcher's analysis in the form of a questionnaire. To formulate the problem, researchers analyze the main material related to temperature and heat to make statements on the instruments created . Indicators , sustainability awareness categories, ESD elements , themes and types of statements are things that need to be considered when compiling this instruments . Bloom's taxonomy operations verbs in the affective and psychomotor domains were used as indicators to create this instruments .

After compiling the instrument, the researchers created a validation sheets for experts and a statement readability test sheets for respondents. The purpose of creating a validation sheets is to assess the suitability of statements with indicators, ESD elements, sustainable knowledge categories, and types of statements so that the instruments can be tested. Another purpose of the readability test sheets is to ensure that the respondent can understand each word in the statement. After that , the completed survey is sent to the validator for validation.

3. Development

The Development Stage is the researcher's stage in making the design (blueprint) a reality. At this stages everything that is needed or that will support the learning process must all be prepared . This stages started with expert validation carried out by 3 physics education lecturers and one of the physics teachers at one of the Bandung city high schools . In the validation process , the validator uses a validation sheets that has been prepared by the researchers . This validation was carried out to assess the content and construct validity for each statement items. At the expert validation stage, validators are asked to provide an assessment of the questionnaire developed based on indicators, ESD aspects, sustainability awareness categories and types of statements as well as provided criticism and suggestions for making improvements so that researchers can perfect the instruments .

Next, the researchers conducted a limited test of the instruments by distributing readability tests to the respondents by distributing questionnaires using Google Forms to class XI high school students in the city of Bandung. After that, the researchers will process and analyze data on the results of the questionnaire assessment obtained from validators and respondents during limited trials to further improve the reference study, pre-writing , draft writing , and the concept of temperature and heat material.

4. Implementation

After the questionnaire has been revised, the next stage is implementation. At the implementation stage, the instruments that have been developed are then improved in such a way that they can be implemented. After the instrument was ready, a wide trial was carried out with a larger number of respondents compared to the limited trial through a large group by distributing questionnaires using Google Form to

various class XI high schools in the city of Bandung. The questionnaire distributed has two levels. The first level is a statement and the second level is the reason for answering the statement. Students involved as respondents must have carried out learning related to temperature and heat materials. Then the instrument developed is tested and the validity and reliability values of the instrument are analyzed and then evaluated.

5. Evaluation

   This stages begins by processing data that has been obtained from extensive trials. After processing , it was found that the questionnaire statements items were valid and had good reliability so that the results could be made regarding the appropriateness of the quality of the instruments that could be used to measure students' sustainability awareness . It was found that the questionnaire statements items were valid and had good reliability.

   The participants involved in this research were representatives of class XI students in several high schools in the city of Bandung. The population used in the research were all class XI students in high schools in the city of Bandung, while the research samples used was 120 class The simple random sampling technique is a technique for taking samples members from a population that is carried out randomly without paying attention to the strata within that population (Sugiyono., 2017).

Data Anlysis

Limited Cova Test Data Collection Technique

At this stage, data collection was obtained from expert assessments of the sustainability awareness instrument that had been developed, testing the readability of the instrument statements to class XI students and testing limited validity and reliability. Expert assessment was carried out to determine the suitability of what was developed with physics learning material on the subject of temperature and heat. In the process of developing this instrument, the experts were appointed to provide assessments were physics expert lecturers and high school physics teachers. Physics expert lecturers are tasked with providing assessments in terms of content of material, relationship of material to ESD aspects and awareness categories, suitability of indicators with the instrument as a whole using a judgment sheet. Assessments related to comments, criticism and suggestions provided by experts are taken into consideration when making revisions and improvements. To test readability, validity and reliability, it was carried out by distributing questionnaires using Googleform to class XI students in one of the high schools in the city of Bandung through physics teachers.

Extensive Trial Data Collection Techniques

Trial data collection wide done with spread instrument existing *sustainability* awareness repaired to student class XI in the city of Bandung with method send *Googleform link* through a physics teacher.

Rash Modeling Analysis

Instrument *sustainability awareness* analyzed use modeling *Rasch* Because can connect between students and items. The success of answering questions correctly depends on the respondent's ability and the level of difficulty of the item. For example, if there are students who are able to do 85% of all the questions correctly, they will definitely have better abilities than other students who are only able to do 60% of all the questions. This statement only shows that the raw data obtained is a type of data that shows ranking and is not linear. Ordinal data does not have equal intervals, resulting in the data needing to be converted

into ratio data for statistical analysis purposes. The ratio used in question is that when a student gets a score of 85% on all questions, the probability of success is 85:15. The existence of this ratio data, *Rasch modeling* can determine the relationship between the level of student ability (*person ability*) and the level of difficulty of items (*items difficulty*) so it can be concluded that the level of success of students with a high level of ability will be able to work on questions with a lower level of difficulty (Bond & Fox , 2007). *Rasch* Modeling assumes that item difficulty is a trait that is influenced by the respondent's answer, and a person's ability is a trait that is influenced by the estimated item difficulty (Linacre, 2011).

Advantages of *Rasch modeling* namely having the ability to make predictions about missing data *based on* systematic response patterns and being able to produce standard *error measurement values* for the instruments used that can increase the accuracy of calculations (Sumintono & Widhiarso , 2014). *Rasch* modeling according to Carvalho et al. (2012) can predict missing data because modeling uses a mathematical model to represent a testing situation, where one person answers a series of items. The more intense a characteristic is in the person, the greater the likelihood of agreement with statements measuring this characteristic. Conversely, the less intense the feature, the less likely it is that the person will agree. *Error* identification is based on the level of capability where according to Rasch (in Sumintono & Widhiarso , 2014, pp. 68-69) says that individuals who have a greater level of ability than other individuals should have a greater chance of answering questions correctly, based on the same principle that items are more difficult causing the individual's opportunity to answer it to be smaller. These advantages are taken into consideration by researchers when using *Rasch modeling.*

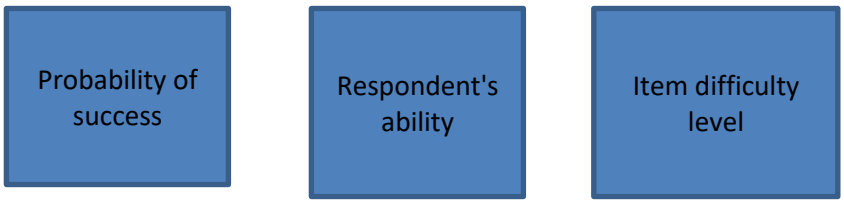
Rasch Modeling for dichotomous data, it combines an algorithm that states the probabilistic expected results of item 'i' and respondent 'n', which is mathematically expressed as follows.

$$P_{ni}(X_{ni} = 1/\beta_n , \delta_i) = \frac{e^{(\beta_n - \delta_i)}}{1+e^{(\beta_n - \delta_i)}}$$

where : $P_{ni}(X_{ni} = 1/\beta_n , \delta_i)$is the probability of respondent n in item i to produce the correct answer (x=1); with the respondent's ability $\beta_n$and item difficulty level δ_i. The equation can be further simplified by inserting a logarithmic function and making it:

$$\log (P_{ni}(X_{ni} = 1/\beta_n , \delta_i)) = \beta_n - \delta_i$$

Based on the equation above , the probability of a success can be written as follows:



(Sumintono & Widhiarso , 2014)

Modeling *Rasch* give information to researcher about data in the form of analysis statistics *The fit* statistics obtained are ideal to describe Students who have high ability provide answer patterns to items according to their level of difficulty . The parameters used are infit and outfit from the middle square (*mean square*) and standardized value (*standardized values*). According to (Widhiarso, 2013), *infit* (*inlier sensitive or informational weighted fit*) measures the sensitivity of response patterns to target items in respondents (persons) or vice versa; while *the outfit* (*outlier sensitive fit*) measures the sensitivity of response patterns to items with a certain level of difficulty for the respondent or vice versa . Quality instrument study determined by validity and reliability from

instrument the . Validity instrument show accuracy measurement in measure what you want to measure or not yet , while reliability shows trust a measurement because of its constancy (Arikunto, 2010).

Instrument *sustainability awareness* in study This using a scale model *likert* consisting of from 4 categories . Where 1 represents very no agree and 4 represents strongly agree . Data obtained from testing limited entered to in *Microsoft Excel* . Then processed use modeling *Rasch* via application *Winstep 3.73* . Tests carried out that is:

1. Validity Test in Rasch Modelling

Rasch modeling analysis. Validity is accuracy tool evaluation For do measurement against concept being assessed. Validity influenced by tools measure (instrument), user tool measure who does measurement and the subject being measured (Sugiyono., 2017). Rasch modeling can detect respondents who are not appropriate in data collection and exclude them if it does not match the existing model. Besides that modeling rasch can also differentiate the ability of respondents between those who are able and those who are not able , so that items can revised or discarded. For want to see the value, click " table output " then click " variable maps " , and to see the numbers per item click "item: measure ". What you see is based on the Outfit value Mean Square (MNSQ), Outfit Z-Standard (ZSTD), and Point Measure Correlation (Pt Mean Corr).

Then , third mark the compared to with criteria (Widhiarso, 2013) as following:

a) Outfit Mean Square (MNSQ) value accepted : 0.5 < MNSQ < 1.5.
b) Outfit Z-Standard (ZSTD) value received : -2.0 < ZSTD < +2.0.
c) Point Measure Correlation Value (Pt Mean Corr): 0.4 < Pt Mean Corr < 0.85.

Statement items instrument sustainability awareness is declared valid at least with fulfill two categories the (Sari, 2014).

2. Reliability Test in Rasch Modeling

Reliability is a measurement process that provides results consistent when done with a measuring instrument it is done in different times . Increasing reliability coefficient big shows a smaller error in measurement, so it can be said that the measuring instrument is more reliable. On the other hand, a smaller reliability coefficient means that the measurement error is greater and the measuring instrument is increasingly unreliable (Azwar, 2012). To determine the reliability of the instrument , researchers used the reliability coefficient from the results of data analysis using the Rasch model with software Winsteps as well as Alpha coefficient Cronbach . The following is the categorization of reliability coefficients in data analysis using the Rasch model and reliability coefficients. Criteria according to(Widhiarso, 2013) as follows:

a. Person Measure, a mean value higher than logit 0.0 indicates student ability is greater than the item difficulty level.
b. Alpha Value Cronbach , the interaction between person and item as a whole. Alpha Criterion Cronbach's is in Table 1 as follows.

Tabel 1. Category Cronbach Alpha Reliability

Logit Value	Interpretation
0.80 < x ≤ 1.00	Very good
0.70< x ≤ 0.80	Good
0.60<x<0.70	Enough
0.50 < x ≤ 0.60	Bad
0.00 ≤ x ≤ 0.50	Bad

c. Pesonality Reliability and Item Realiability Values, person and item reliability values in Rasch modeling have the criteria described in Table 2 as follows.

Tabel 2. Classification Reliability in Data Analysis Using Rasch Model

Logit Value	Interpretation
0.94< x ≤ 1.00	Special
0.91< x ≤ 0.94	Very good
0.81< x ≤ 0.91	Good
0.67< x ≤ 0.81	Enough
0.00 ≤ x ≤ 0.67	Weak

d. The grouping of people and items can be known from the separation value . The greater the separation value , the better the quality of the instrument based on the total number of students.

3. Rating Scale Analysis

Analysis scale ranking done For verify is ranking options used confusing for respondents or No . For get information about scale ranking use analysis modeling rasch can accessed use application *Winstep* with choose table 3.2 Rating (partical credit) scale later see the average observation in the OBSVD AVRGE column . If the logit value is in the option One until end experience increase , interpreted that respondents understand the differences between each alternative answer . Apart from that, you can seen through table *A ndrich Reshold* For test accuracy mark polytomy used . *A ndrich* value *Reshold* which moves in a way in sequence from *none* to negative Then leads to positive show that given options is valid for respondents , however If in table A *ndrich* value *Reshold* No sequentially can interpreted that choice in instrument must simplified . *Sustainability* instrument *awareness* of research This using a *Likert scale* 4 points with Alternative respondents' answers are in the form of:

Tabel 3. Criteria Rating Scale Analysis

Value	Interpretation
SS	Strongly Agree
S	Agree
TS	Disagree
STS	Strongly Disagree

4. Undimensionality Analysis in Rasch Modeling

Unidimensionality instrument is the most important measurement For know and evaluate necessary scale measured , in matter This is measuring sustainability awareness instruments . In app Winstep, got it analyzed through Table 3. Dimensionality Map with method see results measurement raw variance. The un i dimensionality criteria in the Rasch model are presented in Table 3 as follows.

Tabel 4. Criteria Undimensionality

Score	Criteria
>60%	Special
40-60%	Good
20-40%	Enough
≥20%	Minimal
<20%	Bad
<15%	Unexpected Variance

3. RESULTS AND DISCUSSION

Necessary thing Notes in this research include the student's answer choices for each statement along with the student's reasons for choosing the answer choice, whether they strongly disagree, disagree, agree or strongly agree for all statements. Students' reasons are discussed based on the sustainability category awareness in the form of sustainability practice awareness , behavioral and attitude awareness and emotional awareness . The discussion of this is as follows.

A. Sustainability Practice Awarness

Sustainability category statement practice awareness with the most respondents choosing to strongly agree was statement number 3 at 30 %, while the statement with the fewest respondents choosing to strongly agree was statement number 15 at 0 %. Apart from that, there is also a sustainability category statement practice awareness where respondents chose to agree the most was statement number 3 at 63.3 %. Statement number 15 is the statement with the most respondents choosing to strongly disagree at 86.7 % compared to the choice of strongly disagreeing with the sustainability statement. practice awareness another. The statements where respondents dominated the choice of agree were statements number 3 and 1 2.

B. Behavioral and Attitude Awareness

Statements that fall into the behavioral category and attitude awareness are numbers 1, 4, 7, 10, and 13 . Behavioral category statement and attitude awareness with the most respondents choosing to strongly agree was statement number 1 3 amounting to 83.3 %, while the statement with the fewest respondents selecting strongly agree was statement number 4 amounting to 6.7 %. Apart from that, there are also behavioral category statements and attitude awareness where respondents chose to agree the most was statement number 1, amounting to 4 3.3 %. There are statements number 1 , 4, 7, 10, and 13 respondents No there are those who really choose not agree . Statement number 4 includes behavioral statements and attitude awareness , the majority of respondents chose to strongly disagree at 86.7%. There are statements number 1 and 13 respondents No someone chooses No agree . The statement where the majority of respondents chose to agree was statement number 1.

C. Emotional Awareness

Statements that fall into the emotional category awareness are numbers 2, 5, 8, 11, and 14. Emotional category statement awareness The one with the most respondents choosing to strongly agree was statement number 2 at 7 3.3 % while the statement with the fewest respondents choosing to strongly agree was statements number 1 1 and 14 at 3.3 %. Apart from that, there are also emotional category statements awareness where respondents chose to agree the most was statement number 5 at 56.7%. Statement number 5 is an emotional statement awareness of which none of the respondents chose to strongly disagree. Statements where respondents dominated the no choice agree are statements number 6, 11, and 14.

Uni-dimensionality of Sustainability Instruments Awareness

```
    Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)
                                        -- Empirical --    Modeled
Total raw variance in observations   =      56.0 100.0%       100.0%
  Raw variance explained by measures =      41.0  73.2%        72.4%
    Raw variance explained by persons =       8.6  15.4%        15.3%
    Raw Variance explained by items  =      32.4  57.8%        57.1%
  Raw unexplained variance (total)   =      15.0  26.8% 100.0%  27.6%
    Unexplned variance in 1st contrast =      3.9   7.0%  26.2%
    Unexplned variance in 2nd contrast =      2.9   5.1%  19.1%
    Unexplned variance in 3rd contrast =      1.8   3.3%  12.2%
    Unexplned variance in 4th contrast =      1.6   2.9%  10.7%
    Unexplned variance in 5th contrast =      1.2   2.2%   8.2%
```

Figure 1. ***Improved Questionnaire Unidimensionality Measurement Results***

Based on results findings reliability instrument in Figure 1 use Rasch modeling. The results of the unidimensionality test show a raw variance value of 73.2 % which is in the special category based on the unidimensionality criteria in Rasch modeling (Rasch Model). An instrument is said to be reliable if the *raw value variance explained by measure* minimum 20% and *unexplained value variance in the first contras* t does not exceed 15% (Rosli , et al., 2020).

A. Development Stage

```
INPUT: 30 Person  15 Item  REPORTED: 30 Person  15 Item  4 CATS    WINSTEPS 3.73
-------------------------------------------------------------------------------
Person: REAL SEP.: .76  REL.: .37 ... Item: REAL SEP.: 5.65  REL.: .97

         Item STATISTICS:  MEASURE ORDER

-----------------------------------------------------------------------------------------
|ENTRY   TOTAL  TOTAL          MODEL|   INFIT  |  OUTFIT  |PT-MEASURE |EXACT MATCH|      |
|NUMBER  SCORE  COUNT  MEASURE  S.E. |MNSQ  ZSTD|MNSQ  ZSTD|CORR.  EXP.| OBS%  EXP%| Item |
|------------------------------------+----------+----------+-----------+-----------+------|
|    15     36     30    2.76     .41| .92    .0| .87   -.1|  .41   .31| 86.7  82.5| n15  |
|     8     39     30    2.34     .35|1.72   1.8|1.06    .3|  .59   .36| 80.0  72.6| n8   |
|    11     39     30    2.34     .35|1.49   1.3|1.17    .5|  .34   .36| 76.7  72.6| n11  |
|     9     42     30    2.02     .31|1.12    .5|1.08    .3|  .65   .40| 76.7  66.2| n9   |
|     4     47     30    1.61     .27| .72  -1.1| .73   -.9|  .10   .43| 53.3  55.3| n4   |
|    14     47     30    1.61     .27|1.28   1.1|1.23    .8|  .18   .43| 43.3  55.3| n14  |
|     6     56     30    1.04     .24| .83   -.7| .79   -.9|  .70   .46| 43.3  37.4| n6   |
|    12     69     30     .36     .22|1.09    .5|1.17    .8|  .06   .44| 46.7  41.6| n12  |
|     5     96     30   -1.19     .27| .64  -1.4| .66  -1.3|  .42   .35| 50.0  55.6| n5   |
|     3    103     30   -1.77     .31| .46  -2.3| .51  -2.1|  .59   .31| 83.3  57.7| n3   |
|     2    104     30   -1.87     .31| .84   -.5|1.02    .2| -.05   .30| 46.7  57.7| n2   |
|    13    106     30   -2.07     .33| .79   -.7| .84   -.5|  .19   .29| 56.7  59.2| n13  |
|     1    107     30   -2.18     .33| .56  -1.8| .59  -1.5|  .59   .28| 86.7  59.3| n1   |
|    10    107     30   -2.18     .33|1.56   1.8|1.49   1.6|  .30   .28| 60.0  59.3| n10  |
|     7    112     30   -2.84     .40|1.23    .8|1.37   1.0|  .10   .23| 76.7  74.4| n7   |
|------------------------------------+----------+----------+-----------+-----------+------|
| MEAN    74.0   30.0     .00     .31|1.02   -.1| .97   -.1|           | 64.4  60.5|      |
| S.D.    30.1    .0    1.98     .05| .37   1.3| .28   1.0|           | 16.3  11.5|      |
-----------------------------------------------------------------------------------------

▲TABLE 13.3 C:\Users\lenovo\Documents\Book2.prn   ZOU804WS.TXT  Jan 20  3:30 2021
INPUT: 30 Person  15 Item  REPORTED: 30 Person  15 Item  4 CATS    WINSTEPS 3.73
-------------------------------------------------------------------------------
```

Figure 2. Validity Test Results

The dimensionality of the instrument has been corrected after *judgment* can be seen from Figure 2 . The questionnaire has *raw variance* data amounting to From the picture above , the questionnaire has *raw variance* data amounted to 74.5 %. Apart from that, *unexplained variance in the first contrast* (unexplained variance in the first

contrast) of 7.2 %, *unexplained variance in the second contrast* (unexplained variance in the second contrast) was 4.7 %, *unexplained variance in the third contrast* (unexplained variance in the third contrast) was 3.0 %, *unexplained variance in the fourth contrast* (unexplained variance in the fourth contrast) of 2.3 %, *unexplained variance in the fifth contrast* (unexplained variance in the fifth contrast) was 1.9 %. Based on the criteria in Table 3, unidimensionality are in the category special , meaning that the instrument used measures one variable, namely sustainability awareness , without being influenced by other variables.

B. Evaluation Stage

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
1	420	120	-2.47	.18	1.08	.7	.98	-.1	.62	.50	66.7	67.5	n1
2	368	120	-.96	.16	.80	-1.7	.77	-2.0	.75	.57	56.7	61.6	n2
3	404	120	-1.96	.17	.72	-2.4	.66	-2.7	.69	.54	86.7	66.1	n3
4	324	120	.12	.15	.62	-3.5	.62	-3.5	.44	.59	80.0	58.1	n4
5	392	120	-1.61	.17	.83	-1.4	1.09	.7	.23	.55	66.7	64.7	n5
6	328	120	.03	.15	1.64	4.4	1.64	4.4	.81	.59	46.7	58.7	n6
7	424	120	-2.60	.19	.75	-2.1	.62	-2.6	.73	.50	76.7	67.7	n7
8	296	120	.77	.15	.65	-3.2	.65	-3.2	.74	.59	76.7	59.6	n8
9	188	120	3.45	.17	1.50	3.5	1.53	3.4	.34	.57	56.7	65.6	n9
10	420	120	-2.47	.18	.67	-2.8	.58	-3.1	.61	.50	80.0	67.5	n10
11	168	120	4.12	.19	1.41	2.6	1.64	3.1	.10	.54	53.3	71.8	n11
12	228	120	2.38	.16	.70	-2.8	.72	-2.6	.59	.60	66.7	59.9	n12
13	456	120	-3.97	.24	1.09	.6	1.53	1.5	.20	.37	83.3	80.9	n13
14	240	120	2.08	.16	1.00	.0	.97	-.2	.66	.60	73.3	59.2	n14
15	200	120	3.10	.17	1.24	1.9	1.25	1.9	.47	.58	33.3	61.1	n15
MEAN	323.7	120.0	.00	.17	.98	-.4	1.02	-.3			66.9	64.7	
S.D.	94.5	.0	2.47	.02	.32	2.5	.39	2.6			14.5	5.9	

Figure 3. Improved Instrument Validity Test Results

The dimensionality of the instrument has been improved after testing can be seen from Figure 3 . The questionnaire has raw variance data amounted to 73.2 %. Apart from that, unexplained variance in the first contrast (unexplained variance in the first contrast) of 7.0 %, unexplained variance in the second contrast (unexplained variance in the second contrast) was 5.1%, nexplained variance in the third contrast (unexplained variance in the third contrast) was 3.3 %, unexplained variance in the fourth contrast (unexplained variance in the fourth contrast) was 2.9 %, unexplained variance in the fifth contrast (unexplained variance in the fifth contrast) was 2.2 %. Based on the criteria in Table 3, unidimensionality are in the category special , meaning that the instrument used measures one variable, namely sustainability awareness , without being influenced by other variables . Based on the instrument reliability criteria , the instrument has been repaired after judgment has entered the category of a reliable instrument . This is because it is raw variance data is above 20% and unexplained variance in the first contras t does not exceed 15%. Based on this discussion, sustainability instruments The awareness developed is reliable , meaning the instrument developed is capable of measuring sustainability awareness high school student.

Validity of the Sustaianibility Instrument Awareness

A. Development stage

INPUT: 30 Person 15 Item REPORTED: 30 Person 15 Item 4 CATS WINSTEPS 3.73

Person: REAL SEP.: .76 REL.: .37 ... Item: REAL SEP.: 5.65 REL.: .97

Item STATISTICS: MEASURE ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	INFIT ZSTD	OUTFIT MNSQ	OUTFIT ZSTD	PT-MEASURE CORR.	PT-MEASURE EXP.	EXACT MATCH OBS%	EXACT MATCH EXP%	Item
15	36	30	2.76	.41	.92	.0	.87	-.1	.41	.31	86.7	82.5	n15
8	39	30	2.34	.35	1.72	1.8	1.06	.3	.59	.36	80.0	72.6	n8
11	39	30	2.34	.35	1.49	1.3	1.17	.5	.34	.36	76.7	72.6	n11
9	42	30	2.02	.31	1.12	.5	1.08	.3	.65	.40	76.7	66.2	n9
4	47	30	1.61	.27	.72	-1.1	.73	-.9	.10	.43	53.3	55.3	n4
14	47	30	1.61	.27	1.28	1.1	1.23	.8	.18	.43	43.3	55.3	n14
6	56	30	1.04	.24	.83	-.7	.79	-.9	.70	.46	43.3	37.4	n6
12	69	30	.36	.22	1.09	.5	1.17	.8	.06	.44	46.7	41.6	n12
5	96	30	-1.19	.27	.64	-1.4	.66	-1.3	.42	.35	50.0	55.6	n5
3	103	30	-1.77	.31	.46	-2.3	.51	-2.1	.59	.31	83.3	57.7	n3
2	104	30	-1.87	.31	.84	-.5	1.02	.2	-.05	.30	46.7	57.7	n2
13	106	30	-2.07	.33	.79	-.7	.84	-.5	.19	.29	56.7	59.2	n13
1	107	30	-2.18	.33	.56	-1.8	.59	-1.5	.59	.28	86.7	59.3	n1
10	107	30	-2.18	.33	1.56	1.8	1.49	1.6	.30	.28	60.0	59.3	n10
7	112	30	-2.84	.40	1.23	.8	1.37	1.0	.10	.23	76.7	74.4	n7
MEAN	74.0	30.0	.00	.31	1.02	-.1	.97	-.1			64.4	60.5	
S.D.	30.1	.0	1.98	.05	.37	1.3	.28	1.0			16.3	11.5	

TABLE 13.3 C:\Users\lenovo\Documents\Book2.prn ZOU804WS.TXT Jan 20 3:30 2021

INPUT: 30 Person 15 Item REPORTED: 30 Person 15 Item 4 CATS WINSTEPS 3.73

Figure 4. Validity Test Results

Validity of instruments that have been corrected after *judgment* can be known through analysis of Figure 4. Based on the criteria above , all valid questionnaire items. This matter because mark *MNSQ outfits* , *ZSTD outfits* and *Points Measure Correlation* fulfil condition.

B. Evaluation Stage

The validity of the instrument has been improved after testing can be known through analysis of Figure 4. 3 . An instrument is said to be valid if it meets the conditions discussed previously. In the findings above , validity instrument can analyzed based on figure 4.4. According to (Widhiarso, 2013) There is three criteria validity instrument that is:

1. Outfit Mean Square (MNSQ) value received is : 0.5 < MNSQ < 1.5
2. Outfit Z-Standard (ZSTD) value received is : -2.0 < ZSTD < +2.0
3. Point Measure Correlation (Pt Mean Corr) value received is : 0.4 < Pt Measure Corr < 0.85.

If the test items meet at least two criteria above , then item question can used , in other words items is valid. From the statement the there is four instrument items were invalid out of 15 items. Invalid items are found in numbers 6, 9, 11, and 13. No. 6 has an Outfit Mean Square (MNSQ) value of 1.64 where mark the enter into criteria that are not accepted , for Z-Standard Outfit Value (ZSTD) value 4.4 values the No enter into the accepted criteria , and for the Point Measure Correlation (Pt Mean Corr) value of 0.81 the enter into the accepted criteria . So from results analysis Item number 3 is invalid . No. 9 has an Outfit Mean Square (MNSQ) value of 1.53 where mark the enter into criteria that are not accepted , for Z-Standard Outfit Value (ZSTD) value 3.4 values the No enter into the accepted criteria , and for the Point Measure Correlation (Pt Mean Corr) value of 0.34 the No enter into the accepted criteria . So from results analysis The item number 9 is invalid. No. 11 has an Outfit Mean Square (MNSQ) value of 1.64 where mark the No enter into the accepted criteria , for Z-Standard Outfit Value (ZSTD) value 3.1 values the No enter into the accepted criteria , and

for the Point Measure Correlation (Pt Mean Corr) value 0.1 value the No enter into the accepted criteria . So from results analysis The item number 11 is invalid. No. 13 has an Outfit Mean Square (MNSQ) value of 1.53 where mark the No enter into the accepted criteria, for Z-Standard Outfit Value (ZSTD) value 1.2 values the enter into accepted criteria , and for Point Measure Correlation (Pt Mean Corr) value 0.20 value the No enter into the accepted criteria . So from results analysis Item number 13 is invalid.

Tabel 5. Invalid Instrument

No	Statement	Sustainable Aspects	Sustainability Awareness
6	I use a hair dryer to dry my hair	Economics	Sustainability practice awareness
9	I use the AC at home to cool the room temperature	Economics	Sustainability practice awareness
11	I don't like giving information to friends and family about the importance of protecting the ozone layer	Society	Emotional awareness
13	I prefer to turn off the lights during the day as an effort to reduce electricity use	Economics	Behavioral and attitude awareness

Based on the discussion above , there is a change in item validity between instruments which is improved based on *the judgment results* with instruments that are improved based on the results of *judgment* and trials as well as readability tests. Changes occurred in items number 6, 9, 11, and 13 . Items number 6, 9, 11, and 13 were initially valid, then changed to invalid . These changes are influenced by several factors such as: respondents who did not understand the statements conveyed by the researcher because the item difficulty level is not appropriate to the learning material received by students and data patterns that are not evenly distributed. This is because the researchers did not provide learning treatment in class using the ESD approach in temperature and heat material. These factors are also influenced by internal and external factors from the instrument and factors originating from the students concerned. Therefore, an invalid instrument is not suitable for measuring *sustainability awareness* in students.

Reliability of Sustainability Instruments Awareness

Reliability shows that the instruments used in research to obtain information can be trusted as data collection tools and are able to reveal actual information in the field. A measuring instrument can be said to be reliable if the measurements are carried out repeatedly and the results remain consistent.

A. Development Stage

```
INPUT: 30 Person  15 Item  REPORTED: 30 Person  15 Item  4 CATS    WINSTEPS 3.73
-------------------------------------------------------------------------------

     SUMMARY OF 30 MEASURED Person
-------------------------------------------------------------------------------
|          TOTAL                        MODEL          INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      40.5      15.0         .65     .48       .96    -.1   1.02     .1 |
| S.D.       5.4        .0        1.24     .03       .41    1.0    .51    1.1 |
| MAX.      50.0      15.0        2.95     .55      2.24    2.8   2.49    3.2 |
| MIN.      31.0      15.0       -1.50     .46       .48   -1.5    .42   -1.4 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .51 TRUE SD    1.12  SEPARATION  2.19  Person RELIABILITY  .83 |
|MODEL RMSE    .48 TRUE SD    1.14  SEPARATION  2.35  Person RELIABILITY  .85 |
| S.E. OF Person MEAN = .23                                                   |
-------------------------------------------------------------------------------
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .84

     SUMMARY OF 15 MEASURED Item
-------------------------------------------------------------------------------
|          TOTAL                        MODEL          INFIT        OUTFIT    |
|          SCORE     COUNT     MEASURE   ERROR      MNSQ   ZSTD   MNSQ   ZSTD |
|-----------------------------------------------------------------------------|
| MEAN      80.9      30.0         .00     .35       .98    -.1   1.02    -.1 |
| S.D.      23.6        .0        2.47     .04       .32    1.2    .39    1.3 |
| MAX.     114.0      30.0        4.12     .48      1.64    2.3   1.64    2.3 |
| MIN.      42.0      30.0       -3.98     .30       .62   -1.7    .58   -1.7 |
|-----------------------------------------------------------------------------|
| REAL RMSE    .37 TRUE SD    2.44  SEPARATION  6.56  Item   RELIABILITY  .98 |
|MODEL RMSE    .35 TRUE SD    2.44  SEPARATION  6.99  Item   RELIABILITY  .98 |
| S.E. OF Item MEAN = .66                                                     |
-------------------------------------------------------------------------------
```

Gambar 5. Overall Reliability Logit Value of Questionnaires and Respondents

Based on Figure 5, the questionnaire has been revised after *judgment* from experts has a reliability value seen from " *item reliability* " namely 0.98 , which means according to the classification of(Widhiarso, 2013) is that the overall consistency of the questionnaire items is very consistent and has a *Cronbach* reliability value *alpha* of 0.8 4, which means that the interaction between respondents and items is good so that the instrument can be trusted to be used as a data collection tool.

B. Evaluation Stage

The reliability of the instrument has been improved after testing can be seen from Figure 4.6 which shows that the item reliability value of 0.99 is included in the special category, which means that according to the classification of(Widhiarso, 2013) is that the overall consistency of the questionnaire items is very consistent . According to(Widhiarso, 2013) Measuring the overall reliability of the instrument is analyzed from the Cronbach reliability results alpha , namely the interaction between the respondent or person and the item as a whole . Alpha Value Cronbach's 0.84 is included in the very good category, meaning that the interaction between respondents and items is good so that the instrument can be trusted to be used as a data collection tool.

Based on the discussion regarding Figure 4.5 and Figure 4.6 , instruments have different *item reliability values* equal to 0.0 1 . This difference occurred due to improvements in statements and replacement of statements where the improvement in statements succeeded in improving the reliability of the instrument. For *alpha* value *Cronbach* instrument during the *development stage* with *the evaluation stage* No There is the difference in value is 0.84 . *Alpha* value *Cronbach* on the instruments listed in the research of Hassan, et al. (2010) is 0.81 while the *alpha value Cronbach's* for the instrument

developed after testing was 0.8 4 . On the basis of this statement, it can be seen that the reliability of the instruments developed is not much different because the difference is only 0.0 3 . Therefore, it can be said that the instrument developed is in line with the research instrument of Hassan et al., 2010.

Reliability of Respodents

The reliability of respondents measured in this research is the reliability of respondents who filled out *the Googleform* for testing and the reliability of respondents who filled out *the Googleform.* for extensive testing. The trial test had 3 0 high school students as respondents from the same high school, while the broad test had 1 2 0 high school students as respondents from four high schools in the city of Bandung.

A. Development stage

The reliability of respondents who take part in the trial can be seen from the person reliability value. This figure shows that the reliability value for the people involved in the trial is 0.8 3 . Respondent reliability is included in the very good category.

B. Evaluation stage

The reliability of respondents who take the test can be seen from the person reliability value showing that the reliability value of the people involved in the extensive test is 0.8 3, which means very good.

Based on the discussion above, the reliability of respondents during trial and extensive testing has the same value so the reliability category is the same. It can be concluded that the consistency of respondents' answers is strong because they have the same reliability value (Sumintono & Widhiarso, 2014).

Validity of the Rating Scale Sustainability Instrument Awareness

Analysis scale ranking done For verify is ranking options used confusing for respondents or No . For get information about scale ranking use analysis modeling rasch can accessed use application *Winstep* with choose table 3.2 Rating (partical credit) scale later see the average observation in the OBSVD AVRGE column . Sumintono & Widhiarso (2014) state that If logit value that exists in the option One until end experience increase , interpreted that respondents understand the differences between each alternative answer . Apart from that, you can seen through table *A ndrich Reshold* For test accuracy mark polytomy used . *A ndrich* value *Reshold* which moves in a way in sequence from *none* to negative Then leads to positive show that given options is valid for respondents . *Rating* diagnostics *scale* based on the following: (1) *observed average* is used for vertical arrangement (lowest to highest score); (2) MNSQ *outfit* less than 2.0; (3) threshold differences between adjacent categories (seen from the *Andrich column Threshold*) is between 1.0 and 5.0 logit (Kim & Kim, 2020).

A. Development Stage

Rating analysis scale on instruments that have been repaired after judgment can be seen from Figure 4.8 . Based on this picture, in the picture above it can be seen that the average observation starts from logit -2.19 for choice score 1 (i.e. strongly disagree), logit -1.28 for choice score 2 (disagree), logit +1.20 for choice score 3 (agree) and logit +2.01 for choice score 4 (strongly agree). Outfit value MNSQ for score 1 (i.e. strongly disagree) is 1.13 , for score 2 (disagree) is 0.76 , for score 3 (agree) is 0.8 5 and for score 4 (strongly agree) 1.32 . Apart from that, the image above also shows Andrich's value Thresholds.

B. Stage Evaluation

Rating analysis *scale* on instruments that have been repaired after testing can be seen from Figure 4. 9 . Based on this figure, it can be seen that the average observation starts from the logit -2.79 for option score 1 (i.e. strongly disagree), logit -1.46 for choice score 2 (disagree), logit + 1.48 for choice score 3 (agree) and logit + 3.66 for choice score 4 (strongly agree). *Outfit* value MNSQ for score 1 is 1.26 , for score 2 is 1.10 , for score 3 is 0.90 and score 4 is 0.91 . Apart from that, the image above also shows *Andrich's value Thresholds* which moves from NONE then negative and continues towards positive sequentially. Based on diagnostics according to Kim & Kim (2020), *rating scale* used is valid because it is the average of observations on consecutive scores, *the MNSQ outfit value* received , and the difference in *Andrich* scores *Thresholds* for all scores over 1 logit . Based on the views of Sumintono & Widhiarso (2014) , the options given is valid for respondents and respondents understand the differences between each alternative answer.

4. CONCLUSION

From the results research conducted can concluded there is four items were invalid from the 15 item instrument *sustainability awareness* . these items No can used For measure *sustainability awareness of high school* students on the material temperature and heat , the dimensionality of the instrument developed meets the requirements so that the instrument *is sustainable The awareness* developed is reliable , meaning the instrument developed is capable of measuring *sustainability awareness* high school students, the reliability of the items and *people* involved in the research is in the good category , meaning The instrument developed is reliable and the instrument can be used to measure *sustainability awareness* , the rating scale used by the instrument cannot meet the valid criteria, value *observed average* and *andrich threshold* has increased, which means that respondents understand the differences between each alternative answer, the *raw variance value* is 72.4% which is in the special category , which means The instrument used measures one variable, namely *sustainability awareness* , without being influenced by other variables.

REFERENCES

Arikunto, S. 2010. Prosedur Penelitian Suatu Pendekatan Praktik. Jakarta: Rineka Cipta.

Atmaca, A. C., Kiray, S. A., & Pehlivan, M. (2019). Development of a Measurement Tool for Sustainable Development Awareness. *International Journal of Assessment Tools in Education Vol. 6 No. 1*, 80-91.

Azwar, S. (2012). *Reliabilitas dan validitas (4thed). .* Yogyakarta: Pustaka Pelajar.

Carvalho, L. d., Primi, R., & Meyer, G. J. (2012). Application of the Rasch model in measuring personality disorders. *Trends Psychiatry Psychother, 34(2)*, 101-109.

Hassan, A., Noordin, T. A., & Sulaiman, S. (2010). The status on the level of environmental awareness in the concept of sustainable development amongst secondary school students. *Procedia Social and Behavioral Sciences 2*, 1276-1280.

Kim, H., & Kim, G.-Y. (2020). Kim, H., & KiValidation of the 15-item Brief Inventory of Swallowing Ability (BISA-15) for Older Adults Based on a Rasch Analysis. *Kim, H., & Kim, G.-Y. (2020). Validation of the 15-item Brief Inventory of SCommun Sci Disord 2020; 25(2)*, 489-498.

Linacre. (2011). *A User's guide to WINSTEPS Ministeps; Rasch-model Computer Program. Program Manual 3. 73.*

Mulyatiningsih, E. (2012). *Riset Terapan.* Yogyakarta: UNY Press.

Nursadiah, Suyana, I., & Ramalis, T. R. (2018). Profil Sustainability Awareness Siswa Melalui Integrasi ESD Dalam Pembelajaran Berbasis Masalah Pada Topik Energi di SMP . *Prosiding Seminar Nasional Fisika (SINAFI) 2018* , 207-212.

Segara, N. B. (2015). EDUCATION FOR SUSTAINABLE DEVELOPMENT (ESD) SEBUAH UPAYA MEWUJUDKAN KELESTARIAN LINGKUNGAN. *Segara, Nuansa Bayu. (2015). EDUCATION FOR SUSTAINABLE DEVELOPMENT (ESD) SEBUAH UPAYSOSIO DIDAKTIKA: Social Science Education Journal, II*, 22-30.

Sugiyono. (2015). *METODE PENELITIAN PENDIDIKAN (Pendekatan Kuantitatif, Kualitatif, R&D).* Bandung: ALFABETA.

Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D.* Bandung: Alfabeta, CV.

Sumintono, B., & Widhiarso, W. (2014). *Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial .*

Tristananda, P. W. (2018). MEMBUMIKAN EDUCATION FOR SUSTAINABLE DEVELOPMENT (ESD) DI INDONESIA DALAM MENGHADAPI ISU – ISU GLOBAL. *Jurnal Agama dan Budaya*, 42-49.

Widhiarso, B. S. (2013). *Model Rasch untuk Penelitian Ilmu-ilmu Sosial.* Cimahi: Trimom Publishing House.