

Validity and reliability estimation of assessment ability instrument for data literacy on high school physics material

P E Larasati¹, Supahar¹ and D R A Yunanta²

¹ Physics Education, Universitas Negeri Yogyakarta, Sleman, Indonesia

² Physics Education, MAN 3 Sleman, Sleman, Indonesia

Corresponding author: putriedrynal95@gmail.com

Abstract. Assessment ability instrument for data literacy of physics material that's not accordance with characteristics of the assessment, it can make students less understanding of thermodynamics as a material of physics. Therefore, we need an assessment instrument developed according to Sumadi, which consists of multiple choice tests to determine the ability of data literacy in order to find out the quality of the content and empirical validity, the reliability of the assessment instrument. Analysis of the content validity that uses the Aiken V, the trial is analyzed through the MNSQ INFIT value, the estimated reliability is expressed by the Cronbach alpha coefficient, Item Response Theory (IRT) seen through the total information function and Standard Error Measurement (SEM). The results of this study indicate that the instrument for assessing data literacy ability is valid with a V Aiken value > 0.8 and MNSQ INFIT parameter 1.00 ± 0.06 . Estimation of the Cronbach alpha coefficient value is 0,93. The results obtained are based on the total information function curve and the instrument is reliable to measure the ability of students in the range of -3 to +3. Based on the value of the information function, the ability of students is still low.

Keywords: *validity, reliability, assessment ability instrument, thermodynamics.*

1. Introduction

Indonesia is a developing country that has been in the era of the 21st century with various types of demands, especially in the field of education which must be easily accepted by students and can compete internationally and facilitate learners in the 21st century to experience the best learning experience so that they can achieve the learning objectives effectively [1]. The graduation competency is the completeness of students in class of physics, so that they can go to the next level.

Strategies that can be used to assist students in achieving completeness in physics material, namely by providing an assessment to find out how the level of student understanding. Assessment in the realm of education is part of collecting and processing various information related to student achievements during learning [2]. Assessment has basic objectives, one of which is conformity with the learning objectives to be implemented and make students worked in teams to solve real problem [3]. Assessment does not only measure the success of learning, but to see the effectiveness of the success of students [4]. For arrange an assessment that is correspond to characteristics of a good test or assessment, including (1) validity, (2) reliability, (3) objective, and (4) practice [5], [6]. Assessments are arranged in accordance with indicators that are useful for measuring one of the potentials of students such as data literacy in physics material, so that it can be used properly if it meets the criteria / conditions of a good assessment.



Continuity between learning and assessment must be carried out well so that the objectives of learning activities can be realized. The suitability of the strategy in applying assessment to learning, will help educators know the potential possessed by students [7]. Data literacy is one of the variables needed in the industrial revolution era 4.0. Because, data literacy activities in learning physics will train students to be more consistent in a pedagogical manner and easy to understand the learning process and according from PISA data which said that Indonesia from 2000-2015 had been ranked below for literacy skills [8]. This is in accordance with the concept of learning physics that emphasizes systematic, creative, and objective activities and concludes the events that are around and data literacy is included in the ability to formulate and answer data-based questions, use data as needed, represent data and infer data from information obtained [9]. Before teaching how the concept of literacy in learning, an educator should understand several indicators contained in data literacy and data literacy used in learning can be useful for conveying information into knowledge and practice by collecting data, analyzing and interpreting [10]. Table 1 show the synthesis of several articles related to data literacy.

Table 1. The indicators of data literacy.

Data literacy		
Aspects	Sub- Aspects	Indicators
Collecting the Data	Select variables based on data	Students can choose interrelated variables based on the data presented
	Using data based variables	Students can use the variables obtained based on the data presented
Connecting the Data	Differentiating data	Students are able to distinguish data that has been linked based on the object presented.
	Combing data	Students are able to bring together several interrelated data
Analyze Data	Selecting data	Students can analyze data according to the concept
Interpreting Data	Configure data	Students are able to interpret data that is configured into graphical objects and diagrams.
	Interpreting data	Students are able to interpret the variables contained in the graph into data.
Provide Conclusions	Formulate results	Students can formulate conclusions based on data obtained.
	Reviewing results	Students can plan the interpretation of objects obtained based on the conclusions on the data presented.

One of the material physics that requires data literacy capabilities, namely thermodynamics. This is because thermodynamics is a part of physics material with abstract concepts so it is poorly understood and visualized and has a high complexity [11]. Students have a low level of mastery of concepts in thermodynamic material [12]. Another error occurred in the thermodynamic material that is, interpreting the graph and using the terms heat and energy in an inconsistent manner and identifying the effort and heat in the P-V diagram which is not quite right, whereas each thermodynamic process is often interpreted through the graph [13]. Various errors that occur in the application of thermodynamic material in learning, can be reduced by applying appropriate measurement variables.

Based on the explanation related to the background of this study, a problem formulation can be obtained namely how the validity and reliability of the assessment instruments used to measure the

ability of high school students' data literacy on physics subject matter thermodynamics? Researchers hope that from solving the main problem, it can become a foothold for practitioners to better know how to compile an assessment instrument used to measure the ability of high school students' data literacy on physics subject matter thermodynamics.

2. Research method

This study uses the stages of developing learning outcomes tests according to Sumadi here the stages that must be carried out are as follows.

2.1. Development of assessment instrument specifications

This stage is used by researchers to find information related to assessment instruments used in MAN 3 Sleman as a research location. The specifications of the assessment instruments developed must be based on the distribution of material in the 2013 curriculum (K13) and know the needs of the questions used in the thermodynamic material at MAN 3 Sleman.

2.2. Writing assessment instruments

This stage is carried out activities compiling assessment instruments in the form of thermodynamic-related questions that are adjusted to the indicators in the data literacy and in the form of multiple choice questions 12 with 5 answer choices with a score of "1" if true and a score of "0" if false.

2.3. Validity of assessment instruments

The validation stage of the assessment instrument is the validation of the assessment instrument to 5 expert judgments. The validation sheet consists of 17 assessment indicators related to aspects of material, construction and language. The validation activity is used to find out the validity and feasibility of each item to be used in the empirical test. Thus, the results of the validation carried out by 5 expert judgments were then analyzed using the Aiken V formula which revealed the attributes in accordance with the indicators and items in the test. [14], as follows:

$$V = \sum_{i=1}^{c-1} \frac{in_i}{N(c-1)} \quad (1)$$

Information, V is index of Aiken validity, i is category of i , n_i is the score given by the average of i , c is number of ratings / criteria, N is number of rater.

The provisions of the Aiken V index are if the Aiken V index is less than 0.4, then the validity is said to be low, between 0.4-0.8 the validity is said to be moderate, and if more than 0.8 the validity is said to be high, after the scores were attained from the experts, quantitative data were converted into qualitative data in order to identify the product quality [15].

2.4. Testing empirical assessment instrument items

At this stage the activity was tested empirically about the questions by involving 12th grade high school students in MAN 3 Sleman, which consisted of 2 classes with 30 students in each class. Validation is empirically tested using both classical and modern approaches (IRT) to see the validity of each item, the level of difficulty of the item and the function of item information.

2.4.1. Evidence of empirical validity. The model used in this study is the Rasch model with 1 PL. Item items are declared valid if the INFITMNSQ value on each item is in the range of 0.77 to 1.30 [16]. The Rasch model has the condition whether an item has been analyzed or not, which has $INFIT \leq 2$, the level of difficulty of a good item ranges from -2 to +2. Between difficult and easy questions must be balanced. therefore, researchers must choose the level of difficulty items according to the provisions.

2.4.2. Reliability estimation. The reliability estimation of the item literacy ability assessment point is shown through the cronbach alpha coefficient. The items are said to be reliable if the reliability value is more than 0.85 [17]. The higher the value of item reliability the more convincing that the items or

respondent trials are appropriate or suitable. Conversely, the lower the reliability value the fewer items or test respondents did not provide the expected information.

2.4.3. Information function and measurement error standard. In Item Response Theory (IRT), the item information function becomes one of the factors that affect the quality of an instrument. The information function of an instrument will be high if it is composed of items that have a high information function. Therefore, the correctness of the information function is inseparable from measurement error. Measurement errors in IRT are referred to as Standard Error of Measurement (SEM). In IRT, SEM value does not depend on the ability of the respondent, so the instrument information function shown is more accurate [18]. The SEM value is inversely proportional to the root value of the test information function, as the following equation shows:

$$SEM(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (2)$$

Information: $SEM(\theta)$ is standard error of measurement and $I(\theta)$ is information function.

2.5. Administration of final form assessment instruments

The final stage of this research is the development of assessment instruments, namely activities to select questions that are suitable for use and meet the requirements of a valid and reliable assessment instrument so that it can be used at the stage of field testing (implementation). In addition, the final data in this study can be used as a need assessment in advanced research.

3. Results and Discussion

3.1. Results of validation of assessment instruments

At this stage, the assessment instruments were tested for content validity through the Aiken formula V value on each item. the results obtained based on the validator's assessment, will be used as a reference for researchers to select items that are valid and can be used. In accordance with the provisions of the Aiken V index, the results of all content validity by 5 expert ratings on 12 items are declared valid as in table 2 below.

Table 2. Results of item analysis expert judgment assessment instrument.

Item number	Number of Aiken	Decision
1	0,86	Valid
2	0,86	Valid
3	0,867	Valid
4	0,86	Valid
5	0,86	Valid
6	0,867	Valid
7	0,867	Valid
8	0,867	Valid
9	0,86	Valid
10	0,933	Valid
11	0,933	Valid
12	0,933	Valid

The results of the content validity of the 12 items are shown in table 2. These are declared valid so that they are suitable for use in experiments.

3.2. Results of empirical validity

The empirical validity of the valuation instrument can be seen through the value of item compatibility (goodness of fit). Goodness of fit value is indicated by the value of INFIT MNSQ using the help of Quest software, so that the average and standard deviation results are 1,00 and 0,06. The average value (mean) is valid because it is in the range of 0.77 to 1.30. Figure 1 shows the distribution of MNSQ INFIT values for each item in the Fit Model, as follows.

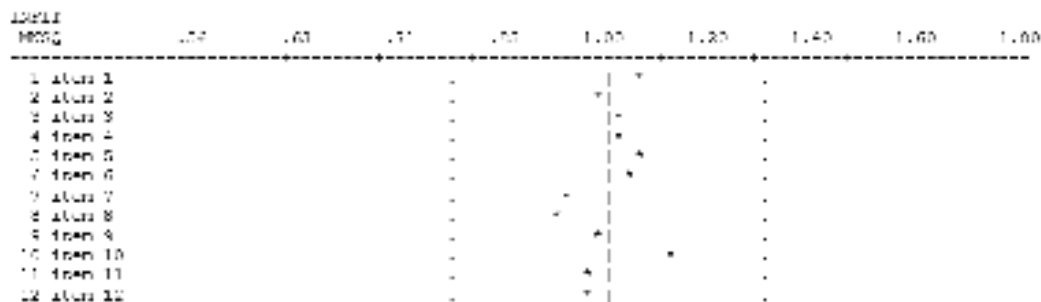


Figure 1. Distribution of MNSQ INFIT values for each item in the FIT model.

The level of difficulty on each item can be seen in the difficulty value (item difficulty level) in the software quest output shown in figure 2 as follows.

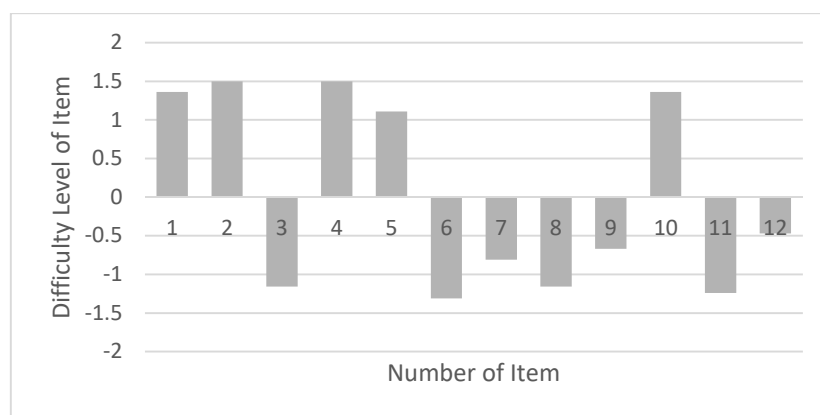


Figure 2. Level of difficulty of item problems in valuation instruments.

Figure 2 is arranged based on the data in the quest software output related to the level of difficulty of the questions given. Thus, Figure 2 shows the results of the distribution of questions with difficulty ranging from easy questions to difficult questions. Items 1, 2, 4, 5, and 10 have difficulty levels approaching +2 which means that the question is difficult. While items 3, 6, 7, 8, 9, 11, and 12 have difficulty levels approaching -2 which means that the questions are easy.

3.3. Results about item reliability

The reliability estimation of the item literacy ability is indicated by the alpha cronbach coefficient of 0.93, so the reliability estimate is reliable because the reliability value is more than 0.85.

3.4. Results information function and standard measurement error

The reliability analysis in figure 3 is still influenced by classical theory, and therefore an analysis based on item response theory (IRT) is also performed. The results of the analysis of the response theory in the form of graphs shown in figure 3, as follows.

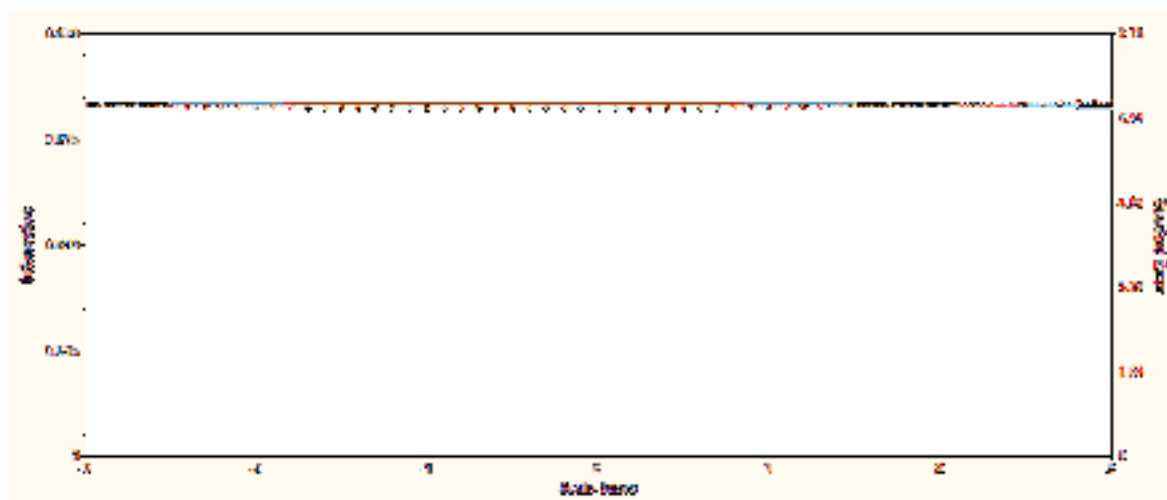


Figure 3. Total information function and standard error measurement (SEM).

Figure 3 above shows that the grading test item has a total information function of 0.016, with a Standard Error Measurement (SEM) of 6.55. Therefore, the instrument for assessing the ability of reliable data literacy to measure the ability of high school students can be known from the Parscale program output at intervals of -3 to +3 can be seen in figure 4, as follows.

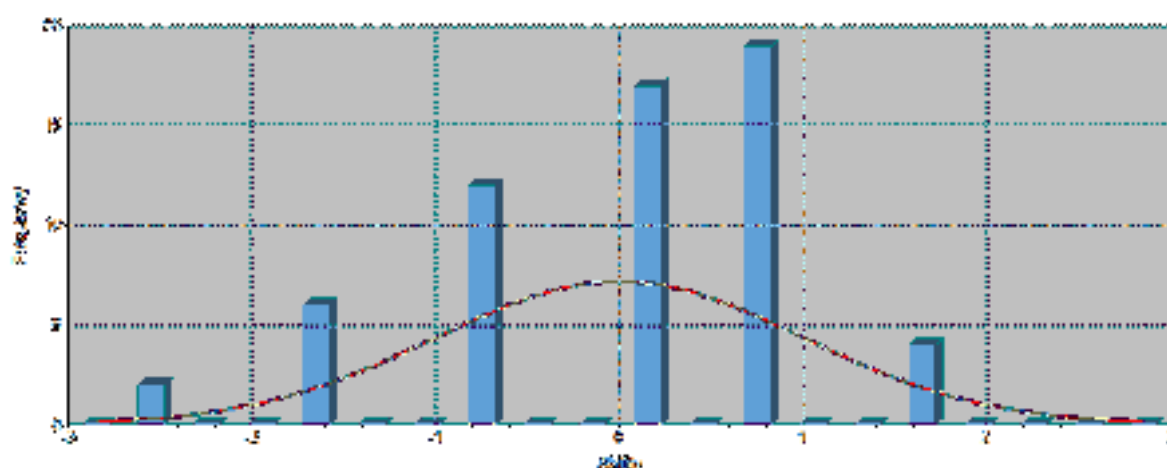


Figure 4. Histogram level of students' literacy data capabilities on thermodynamic materials.

The level of students' ability in figure 4 illustrates that the number of students with low or near-3 abilities is greater than the number of students with moderate ability or more than 0 (close to +3). The results of the analysis on the worthiness of the instrument for evaluating the ability of data literacy can be used to determine the ability of students. Instruments that can be used to see the ability of students, must be in accordance with the terms and conditions of the validity, the reliability of an instrument [18]. One of the results of research that has been done is the benefit of developing tested instruments, namely valid assessment instruments in content validation and empirical validation, one of which is able to measure physics problem solving skills. [19]. This, in accordance with the results of previous studies related to literacy with a valid assessment instrument, is reliable so that it can be used in construct testing [20]. The ability of literacy data can be adequate students to know how to understand the problem of physics material [21]. Based on the results of previous studies, states that the development of valid and reliable assessment instruments can be used to determine students' data literacy abilities.

4. Conclusion

The instrument for evaluating data literacy capabilities is declared valid and is suitable for use. The instrument of assessment is in accordance with the values of the MNSQ INFIT Rasch model. The estimated reliability of the Cronbach alpha coefficient is greater than 0.85. The results obtained are based on the total information function curve, the instrument is reliable to measure the ability of students in the range of -3 to +3. Based on these graphs, it is known that students with medium ability are more than students with high ability. Therefore, it can be concluded that the data literacy ability of students on thermodynamic material is still low so it needs to be followed up in the form of research related to the development of diagnostic assessments to find out the errors experienced by students related to data literacy on thermodynamic material.

References

- [1] Herbert, Putro, B L Putra R R J and Fitriasari N S 2019 *Proc. Mathematics, Science, and Computer Seminar International (Bandung)* vol 1280 (Bristol: IOP Publishing) p 1-7 <http://doi.org/10.1088/1742-6596/1280/3/032032>
- [2] Leighton J P and Rupp A A 2017 *The handbook of cognition and assessment* (United Kingdom: Wiley Handbooks in Education)
- [3] Suastra I W, Ristiati N P Adnyana P P B and Kanca N 2019 *Seminar Nasional Fisika (Surabaya)* vol 1171 (Bristol: IOP Publishing) <http://doi.org/10.1088/1742-6596/1171/1/012027>
- [4] Larson J and Marsh J 2015 *Making literacy real theories and practices for learning and teaching.* (United Kingdom)
- [5] Wardono and Mariani S 2018 *Proc. International Conference on Mathematics, Science and Education (Semarang)* vol 983 (Bristol: IOP Publishing) <https://doi.org/10.1088/1742-6596/983/1/012107>
- [6] Barone D M and Mallette M H 2013 *Best practices in early literacy instruction* (New York: The Guildford Press)
- [7] Putro B L, Waslaluddin Putra R R J and Rahman 2019 *Mathematics science and computer science education international seminar 2018 (Bandung)* vol 1280 (Bristol: IOP Publishing) p 1–7 <https://doi.org/10.1088/1742-6596/1280/3/032034>
- [8] Ovan Waluya S B and Nugroho S E 2018 *Proc. International Conference on Mathematics, Science and Education (Semarang)* vol 983 (Bristol: IOP Publishing) <http://dx.doi.org/10.1088/1742-6596/983/1/012151>
- [9] Hooft M, Swan K, Cook D and Standford T 2012 *Middle Grades Research Journal* **7** 19-33
- [10] Reeves T D and Honig S L 2015 *Teaching and Teacher Education.* <http://dx.doi.org/10.1016/j.tate.2015.05.007>
- [11] Jamuri Kosim and Doyan A 2015 *Jurnal Penelitian Pendidikan IPA* **6** <http://dx.doi.org/10.29303/jppipa.v1i1.11>
- [12] Adila A, Sutopo and Wartono 2018 *Jurnal Pendidikan* **3** <http://dx.doi.org/10.17977/jptpp.v3i10.11629>
- [13] Goldt S and Seifert U 2017 *New Journal of Physics* **9** 1-20 <http://dx.doi.org/10.1088/1367-2630/aa89ff>
- [14] Syarif A N and Kuswanto H 2019 *Proc. The 2nd Annual International Conference on Mathematics and Science Education (Surakarta)* vol 1227 (Bristol: IOP Publishing) <https://doi.org/10.1088/1742-6596/1227/1/012030>
- [15] Ikhsanuddin and Subali B 2019 *Proc. International Conference on Research, Implementation and Education of Mathematics and Science (Yogyakarta)* vol 1097 (Bristol: IOP Publishing) p 1-9 <https://doi.org/10.1088/1742-6596/1097/1/012039>
- [16] Hidayat S and Fadillah E N 2019 *Proc. The International Seminar on Bioscience and Biological Education (Kuala Lumpur)* vol 1241 (Bristol: IOP Publishing) p 8 <https://doi.org/10.1088/1742-6596/1241/1/012040>
- [17] Bashooir K and Supahar 2018 *Jurnal penelitian dan evaluasi pendidikan* **22**

- <http://dx.doi.org/10.21831/pep.v22i2.19590>
- [18] Erfianti L, Istiyono E and Kuswanto H 2019 *International journal of educational research Riview* **4** 320-9 <http://dx.doi.org/10.24331/ijere.573863>
- [19] Erwin, Rustaman N Y, Firman H and Ramalis T R 2019 *Jurnal ilmiah multi sciences* **11** <https://doi.org/10.30599/jti.v11i2.494>
- [20] Batong J S T and Wilujeng I 2018 *Proc. International Conference on Research, Implementation and Education of Mathematics and Science (Yogyakarta)* vol 1097 (Bristol: IOP Publishing) <http://dx.doi.org/10.1088/1742-6596/1097/1/012021>
- [21] Ridho S, Aminah N S and Supriyanto A 2018 *Jurnal penelitian dan pengembangan pendidikan fisika* **5** <http://dx.doi.org/10.21009/1.04201>