Development of assessment instruments to measure problem solving skills in senior high school

A T Astuti[1], Supahar[1], Mundilarto[1] and E Istiyono[1]

[1] Physic Education, Universitas Negeri Yogyakarta, Sleman, Indonesia

Corresponding author: ayut.astuti@gmail.com

developed to measure students' problem solving skills, (2) describe the achievement of students' problem solving skills in SMA Negeri 6 Yogyakarta. The research design used is modification development research of Wilson models and Oriondo and Antonio models. The instrument consisted of 16 essay item that had been validated by measurement experts and practitioners. The sample consisted of 295 students in three high schools in Yogyakarta with high, low, and medium grades. The polytomus data were analyzed according to the partial credit model (PCM) 1 PL by Quest and Parscale application. Research result show that: (1) the developed test instrument in major of Work and Energy is qualified to measure students' problem solving skills because it is proven to be valid and reliable, (2) the problem solving skills of students in SMA Negeri 6 Yogyakarta in major Work and Energy classified in high category with ranges of values θ between -1.27 to 2.81 and mean 0.82.
Keywords: *problem solving skill, assessment, instrument*

1. Introduction

Evaluation and learning are two sides of a coin that cannot be separated [1]. This is in accordance with Ministerial Regulation No. 23 of 2016 concerning Educational Assessment Standards which states that one of the assessment principles is integrated, which means that assessment is an inseparable component of learning activities so that the assessment conducted can be used as feedback, directing learning, and evaluating students and learning that has been done. The minister of education and culture's regulation states the assessment of learning outcomes by educators is the process of gathering information / data about student learning outcomes in aspects of attitudes, aspects of knowledge, and aspects of skills that are carried out in a planned and systematically to monitor the process, learning progress, and improvement of learning outcomes through the assignment and evaluation of results [2]. The methods used to collect educational data define in measurable terms what teachers should teach and what students should learn [3].

Observations and interviews with teachers in SMA N 6 Yogyakarta, SMA N 9 Yogyakarta, SMA N 1 Teladan Yogyakarta, and SMA N 1 Sedayu shows that not many teachers who understand how to develop and use instruments assessment of cognitive corresponding to the dimensions of knowledge and cognitive level to be measured. This causes many assessment activities not found in accordance with the rules of good preparation of assessment instruments. The discrepancy can cause ineffectiveness in the assessment or evaluation activities carried out because the instruments applied are not valid measuring student learning outcomes.

Assessment has an important role to create the 21st century generation, especially assesment of problem solving skills. Problem solving skills need to be possessed by students because in the 21st century students are required to have variety skills, especially thinking skills. ATCS21 divides 21st century skills into 4 groups consisting of (a) living in the world; (b) ways of thinking; (c) tools for working; (d) ways of working. Ways of thingking is a group of thinking skills. Blinkey *et al.* divide the way of thinking into 3 skills namely (1) creative and innovative; (2) think critically, solve problems and make decisions; (3) learning with metacognitive abilities [4].

The results of observations and interviews in several schools in DIY show that there are not many cognitive assessment instruments available that are used to measure students' problem solving skills that have been proven to be valid and reliable. In addition, assessment instruments for problem solving abilities that have been qualified are also still rarely applied to measure students' problem solving abilities in learning. The results of observations and interviews also show that there are still many students who have low problem solving abilities. This can be seen from the lack of students' ability in completing practice questions during learning and in doing exam questions.

Based on the explanation above, it is necessary to have research to develop a cognitive assessment instrument that is focused on measuring students' problem solving abilities on the subject matter of Work and Energy. Therefore, this research aimed to: (1) determine the qualification of cognitive assessment that developed to measure students' problem solving skills, (2) describe the achievement of students' problem solving skills in SMA Negeri 6 Yogyakarta.

2. Research method

2.1 Research method

The research design used in this study was the combination of Wilson models and Oriondo and Antonio models developed by Istiyono. The combination of these two models resulted in a development stage consists of test design, examination of the test instrument, and test assembly [5]. The research procedure is presented in figure 1.

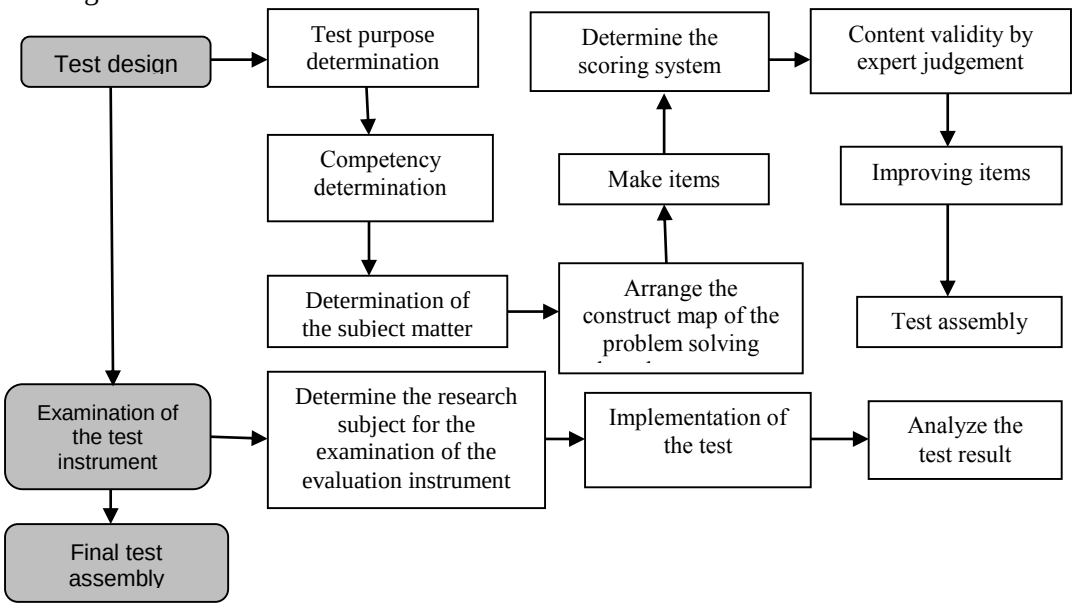

Figure 1. Development procedure

2.2 Data analysis

In this study, the content validity was analyzed using the content validity index from Aiken or Aiken V. Aiken formula is shown by the following equation [6]:

$$V = \frac{s}{[n(c-1)]} \quad (1)$$

Which, V= coefficient or Aiken Validity, n = number of assessor, s = r – lo, lo = the lowest validity score, c = the highest validity score, r = the given scores from the assessor

The test result was polytomous data with four categories. Item analysis was carried out according to Partial Credit Model (PCM) 1 PL using Quest and Parscale program. The Quest program is used to determine reliability of estimate, goodness of fit *(INFIT MNSQ),* and the level of difficulty of the test item, and Parscale program is used to determine information function and standard error of measurement (SEM) and also items ICC curve.

2.3 Research subject

The subject that used in this research is 10th grade of science students of senior high school in Special Region of Yogyakarta. The sample is determined by purposive sampling. Purposive sampling was used to obtain samples consisting of high, medium and low school groups in order to obtain logistical curve. The sample in this research consists of 295 students in SMA A, SMA B, and SMA C, the sample distribution are shown in Table 1. The selection of schools as research subjects is determined by National Examination result in 2016/2017 and 2017/2018.

Table 1. Research subject

School	NE result in 2016/2017	NE result in 2017/2018	Category	Distribution
SMA A	80,65	79,48	High	4 classes
SMA B	75,26	74,45	Medium	3 classes
SMA C	53,58	56,43	Low	3 classes

3. Results and Discussion

3.1 Instrumen development

The instrument test had a total of 16 items. The instrument consists of 2 test packages with total of 9 items for each package including 2 items as anchors. The development of this test is aimed at the physics matter of work and energy. The indicator description of item test are shown in table 4 bellow.

Table 2. Indicators of problem solving ability

Problem solving stages	Indicators	Subindicators
Problem identification	Students identify problems based on basic concepts, make a list of known quantities, determine the quantities in question, and can rewrite problems in different forms (citing problems, drawing diagrams or graphs about problems).	Visualization Identification
Plan the solution	Students can identify concepts, principles, rules, formulas and physical laws related to problems and can write them down in equations systematically.	Plan Formulate
Implement plan's solution	Students can use equations, substitute values, and carry out mathematical operations to find solutions	Correlate Apply Analyze
Evaluate the solution or make answer conclusion	Students can check suitability with the concept, evaluate units on the answers, and draw conclusions from the results obtained.	Check Rate/ criticize

The items distribution is presented on table 3 The assessment instrument developed was validated by 3 expert judgments consisting of physical education experts and physical assessment experts and practitioners.

Table 3. Item distribution

Problem Solving Aspects	Problem Solving Subaspects	Subtopic	Number of Item
Identification	Visualization	Work concept	1
	Identification	Kinetic energy	2
Plan	Plan	Potential energy	3
	Formulate	Mechanical energy	4
	Correlate	Work and kinetic energy relation	5
Implementation	Apply	Work and potential energy relation	6
	Analyze	The law of conservation of	7
Evaluation	Check	mechanical energy	8
	Rate /critizise	Power	9

3.2 Content validity

The results of the analysis stated that the items had an Aiken index of V between 0.83 and 1.00. Based on the results obtained from the analysis using the Aiken V Index, in order to be said to be valid, the Aiken index must have a value with a lower limit of 0.037 and an upper limit of 1.00 for the criteria of three validator with three rating scales [7]. However, according to Sireci and Geisinger (1995) a validity index of 0.7 is still acceptable and belongs to the satisfactory category [8]. Based on the content validation analysis using the Aiken index each item on the developed item instrument was declared valid.

3.3 Empirical validity (fit model)

Validity is empirically proven by goodness of fit in the partial credit model (PCM). The test fit with the model if the average INFIT MNSQ is around 1.0 and the standard deviation is 0.0 or the average INFIT T is close to 0.00 and the standard deviation is 1.0 [5]. Based on data analysis, the average value and standard deviation of INFIT MNSQ are 1.00 ± 0.09 and 1.00 ± 0.20 , so that the average INFIT MNSQ is around 1.0 and the standard deviation is around 0.0, then the test fit with PCM 1 PL. This means that the instrument is empirically valid. The validity of this test is supported by all items having an INFIT MNSQ value of 0.83 to 1.15 which is located between the item acceptance limits using the INFIT MNSQ or fit according to the model (between 0.77 to 1.30) meaning all items developed fit entirely against the model.

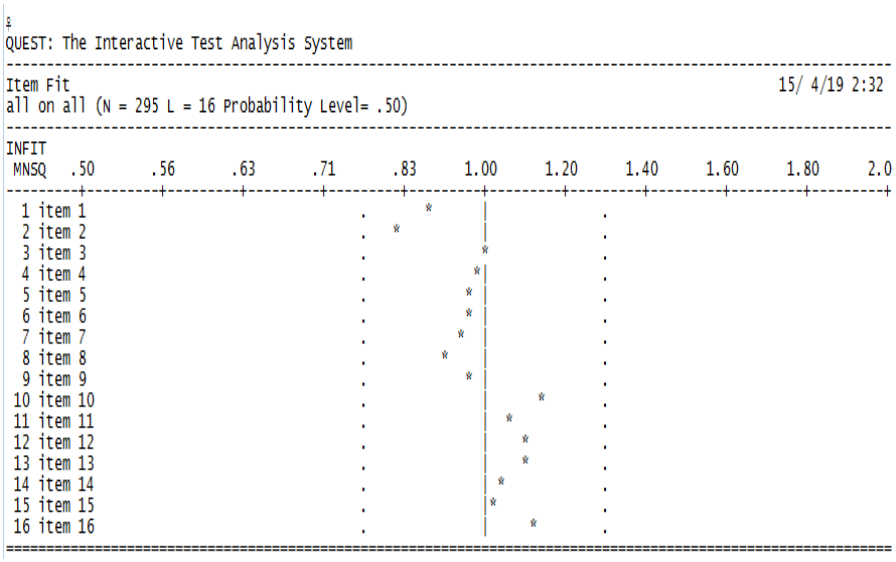

Figure 2. INFIT MNSQ of item developed

3.4 Item difficulty

According to Hambleton and Swaminathan, the level of difficulty (b) of the item is good if it has the item's difficulty index between -2.00 to 2.00 [9]. Items with a difficulty level of -2.00 indicate that the item is very easy, while a difficulty level of 2.00 means the item is very difficult. The following is a graph of the level of difficulty for each item.

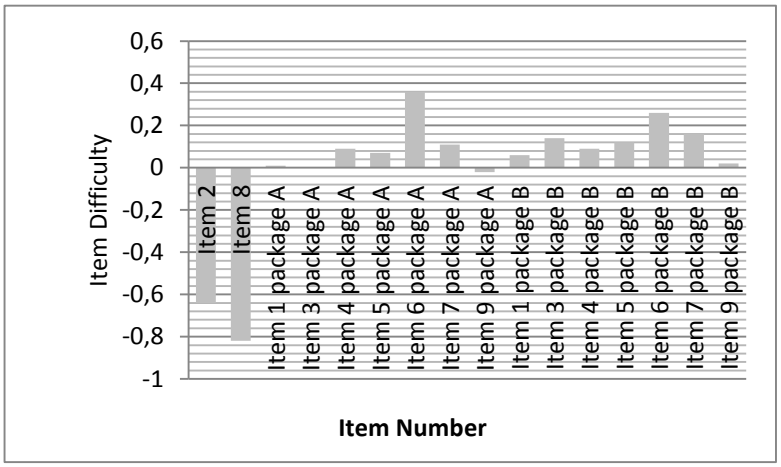

Figure 3. Item difficulty diagram

Thus, in terms of the level of difficulty of the items and their stability, this instrument is included in good categories.

3.5 ICC curve

The curve explains the ability to be able to work on problems in each category. Based on the analysis obtained 16 item characteristic curves (ICC). The Figure below presents an example of ICC for item 12 (item 4 package B), which can be explained that: (a) score 1 (category 1) is mostly obtained by students with very low abilities ($\theta = -3$), (b) score 2 (category 2) mostly obtained by students with low ability ($\theta = -1.5$), (c) score 3 (category 3) mostly obtained by students with high ability ($\theta = 0$), and (d) score 4 (category 4) most of them were obtained by students with very high abilities ($\theta = 3$). Difficulty levels from small to large in order of categories 1, 2, 3, and 4.

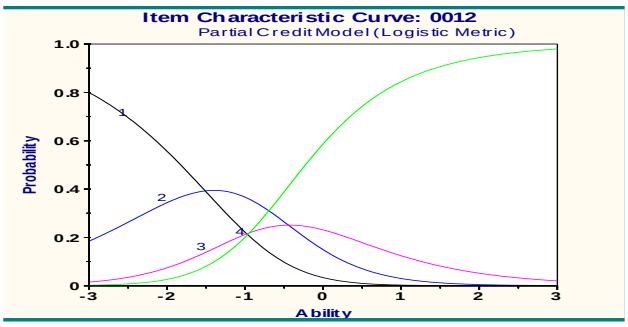

Figure 4. Item ICC Curve 4 package B

Based on the curve, can be concluded that the test scores obtained by students for each item will be in accordance with their level of ability.

3.6 Information function and SEM

Test reliability analyzed by the Quest program obtained a Summary of item estimates or can also be referred as Internal Consistency of 0.64, based on Arikunto (2013), it was concluded that the reliability of the test instruments was categorized high so that it could be used to measure students ability. The reliability of assessment instruments is also determined on an IRT basis using the total Information function and SEM curves. To get the information and SEM functions, the Parscale program is used.

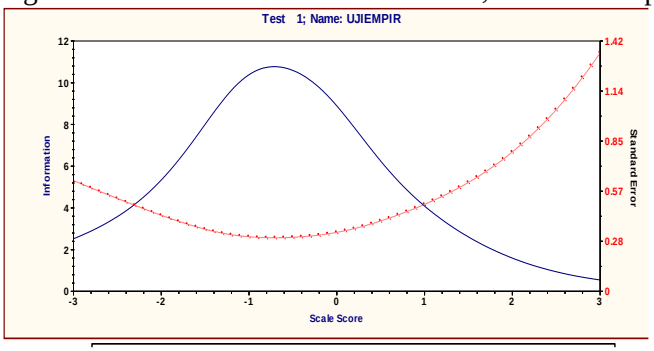

Figure 5. Information function and SEM curves of the test

Based on the result of analysis using the PARSCALE program, it was obtained information fiction and standard error of measurement (SEM). Figure 5 shows that the test is suitable for the students who have the ability (θ) between – 2,3 to 1 (– 2,3 < θ < 1).

3.7 Estimation of students' problem solving abilities in SMA 6 Yogyakarta in major of work and energy

The test instruments that proven to be valid and reliable then used to measure the problem solving ability of class X in Work and Energy major at SMA N 6 Yogyakarta. The data of students' test results are analysed by the Quest program. Estimation on the Quest output is still in the log-add unit, which is students ability minus the difficulty level of the item. Figure bellow show the problem solving abilities mapping for 81 students in SMA 6 Yogyakarta.

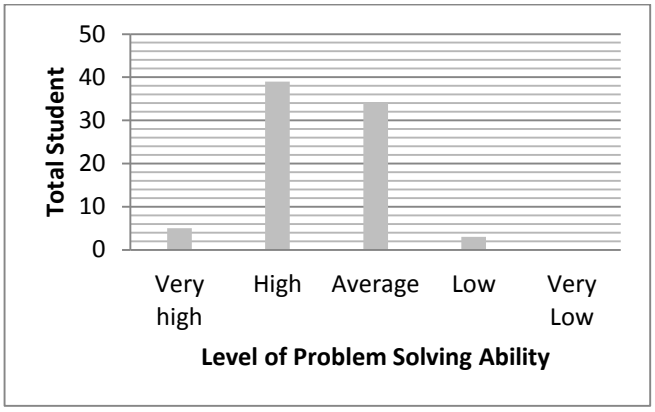

Figure 6. Problem solving ability mapping chart

The distribution of class X problem solving abilities at SMA N 6 Yogyakarta in Work and Energy majorl can be seen in the following table.

Table 7. Distribution of problem solving ability in SMA 6 Yogyakarta

Logit Skor	Criteria	N students	Percentage (%)
$2 < \theta$	Very high	5	6
$0{,}67 < \theta \leq 2$	High	39	48
$-0{,}67 < \theta \leq 0{,}67$	Average	34	42
$-2 < \theta \leq -0{,}67$	Low	3	4
$\theta < -2$	Very low	0	0
Total		81	100
Logit mean		0,82 (high)	

The range of values θ is between -1.27 to 2.81 with a mean of 0.82. From the results of this analysis above, it can be concluded that the problem solving ability of students in SMA N 6 Yogyakarta is included in the high category with an average logit of 0.82 because based on the analysis of the conversion of quantitative values into qualitative values [10] logit score range of $0.67 < \theta \leq 2$ including the high category.

4. Conclusions

The developed test instrument in major of Work and Energy is qualified to measure students' problem solving skills because it is proven to be valid and reliable. The problem solving skills of students in SMA Negeri 6 Yogyakarta in major Work and Energy classified in high category with ranges of values θ between -1.27 to 2.81 and mean 0.82.

References

[1] Temiz K, Tasar F, and Tan M 2006 *International Education Journal* **7** 1007-102
[2] Permendikbud 2015 *Peraturan PERMENDIKBUD NO. 53 Tahun 2015 Tentang Penilaian Hasil Belajar Oleh Pendidik Dan Satuan Pendidikan* (Jakarta: Kemdikbud)
[3] National Research Council 1996 *National Science Education Standards* (Washington DC: National Academy Press)
[4] Binkley M 2012 *Defining twenty-first century skills* In Griffin p, McGaw B, and Care E, Assessment and Teaching of 21st Century Skills p 17-66 (Dordrecht: Springer) http://dx.doi.org/10.1007/978-94-007-2324-5_2
[5] Istiyono E, Mardapi D and Suparno 2014 *Jurnal Penelitian dan Evaluasi Pendidikan* **18** 1-12
[6] Aiken L 1985 *Educational and Psychological Measurement* **45** 131-142
[7] Aiken L 1985 *Educational and Psychological Measurement* **45** 131-142.

[8] Sireci S G and Geisinger K F 1995 *Applied Psychological Measurement* **19** 241-255.
[9] Hambleton R and Swaminathan H 1985 *Item response theory: Principles and application* (Boston: Kluwer)
[10] Azwar S 1998 *Metode penelitian* (Yogyakarta: Putra Pelajar)