Development of learning instrument to measure student learning outcomes on integrated science and Islamic studies

J B Poernomo¹²*, Wiyanto¹, Supartono¹ and A Rusilowati¹

¹ Postgraduated Program of Semarang State University, Semarang, Indonesia
² Department of Physic Education of UIN Walisongo Semarang, Indonesia

*Corresponding author: po3rnomoextra@gmail.com

Abstract. The purpose of this study is to develop a learning outcome measuring instrument of the basic concept of integrated science and Islamic study in accordance with the stages of test development based on the competency standards of graduates. This research stage refers to research development. The research data was collected by using test and non-test. The data was obtained from the students answers to lecture material tests, questionnaires and interviews with the lecturers of Basic Science Concepts subject and the students of Madrasah Ibtidaiyah Teachers-Training Program in the third semester. The sampling technique uses purposive sampling and using qualitative data analysis include language, material conformity, construction, objectivity and systematics. Quantitative data analysis was performed using statistical test. The results of the study are a description of procedural processes and product development which then developed as a means of measuring learning outcomes. The stages taken to develop the learning test instrument are to determine the test objectives, preparing the test specifications, writing the questions, performing the qualitative analysis by 3 expert evaluators then doing revision I. testing the practicality with 4 lecturers then doing revision II. Simulating the use of products in small classes then doing revision III. testing product in one class, item analysis, and doing revision IV. Based on a study of experts measuring results of learning integrated science of Islamic Religion has a very good content validity and high reliability of 0.91. The final result of this research is a measuring instrument of learning result in cognitive domain. This result consists of 31 multiple choice questions test which categorized as valid items and can be implemented.

1. Introduction
The success of education can be measured through three components: curriculum, learning process and assessment activities. All three components have the same position and cannot be separated from each other. Besides a reliable curriculum, the correct learning process still stands in need of a good, directed and planned assessment system [1].

The assessment of college learning outcomes was conducted based on the Standards of Higher Education assessment which refers to aspects of cognitive, affective and psychomotor. In addition, some basic principles must be considered. The principles underlying the assessment of learning outcomes are objective, accountable, educational, authentic, and transparent [2].

Nowadays, there are many learning outcomes that have not met qualifications as a good standardized measurement, especially in the Madrasah Ibtidaiyah Teachers-Training Program. One of the causes is the ability of the lecturers in making test instruments are still lacking so that the
assessment and measurement process has not been as expected before. The lecturers have not fully analyzed the item, so the question instrument used so far has not shown the good quality of the questions used. Therefore, it is necessary to learn the result of learning instrument through the stages of test development, especially, in Islamic universities, which integrate science and Islam study.

The quality of the learning instrument must meet these requirements: content, construction and language, objectivity and systematic. In addition, the learning result instrument must also fulfill the validity through empirical observation. The outcomes of the learning instruments must be valid, reliable and free from bias elements [4].

The response theory of the item is able to explain the characteristics of the test that can be related to the ability measured by the learning instrument. Response items can be measured dichotomous and polytomous. The dichotomous data use quantitative analysis tests to be appropriate so that the test is more reliable and valid than the classical test theory [5].

Development of learning result instruments should follow the pattern of stages systematically can generate maximum results. With the development stage of the test and analysis of the items used well, will be obtained learning outcomes that can measure and recognize the ability of learners [6].

Based on the previous background, the purpose of this research is to develop a learning instrument of integrated science and Islamic study in the lecture of the Basic Concept of Science. This research run in Madrasah Ibtidaiyah Teacher-Training Program based on the expected outcomes. The results of this development research are to obtain the cognitive aspect learning result of the lecture of the Basic Concept of Science in the form of multiple-choice questions. These questions consisting of 5 prepared possible answers based on the material of Basic Concept of Science. These questions have been through semester lecture plan as good question instrument.

2. Method
This research is research development. The research model used in this research as the reference for the implementation of research is the development stage of R & D Model by Sugiyono [7]. Activities taken on in this research include product design, design validation, design revisions, product testing, product revisions, experimental use, product revisions and mass production. This study was conducted only covering the stages of product design until product revision after product trial. The last three stages of trial use, product revision and mass production will take place at the next research.

The study was conducted on the students of Madrasah Ibtidaiyah Teachers-Training Program at the Faculty of Tarbiyah and Teacher Training UIN Walisongo Semarang. The subject of this research is the student of Madrasah Ibtidaiyah Teacher-Training Program in the third semester as the subject of simulation and product trial in trial class. There are three experts as expert evaluators who validated the created test instrument. The experts take role as validator of the content substance, construction, language, objectivity and systematic instrument of student’s integrated science learning outcome.

The result of cognitive aspect instrument that developed in this research is multiple choice questions with 5 possible answers in accordance to the semester lecture plan of integrated science and Islamic study. Then the results of this study will be obtained through focus group discussion (FGD). While the data collection technique done in this research is obtained through the technical test that given to the students using the test sheet of knowledge test instruments, as well as non-test techniques conducted to students and lecturers using questionnaires and interviews.

The data analysis technique is done qualitatively based on the results of design validation by experts in the form of criticism, suggestion and research result as a whole. While quantitative data analysis using result of statistical analysis. Qualitative analysis of the data in the form of validity score of the cognitive domain based on criteria [8] as shown on the Table 1.
## 3. Results and Discussion

### 3.1. Development process on integrated learning instrument

The cognitive learning instrument was developed in the form of multiple-choice questions with five possible answers based on competency standards developed in the semester lecture plan [9]. The process of developing cognitive test instruments is developed according to the stages of research and development.

#### 3.1.1 Product Design

Product design of cognitive test instruments carried out according to test development. The initial step of test development is done by compiling the test specification of knowledge and assembly test questions. The structure of the test cognitive instrument specification consists of: (i) Determining the test objectives: This activity aims to measure the ability of students’ knowledge according to the cognitive level, (ii) Prepare the lattice: The lattice grid used in this study is based on the competency standards of the graduates, (iii) Selecting the cognitive test instrument: The choice on the objective test is in multiple choice with five possible answers, (iv) Determining the length of the Test Instrument: The developed test is 40 questions with a 90-minute time allocation. The composition of the questions is done based on the arrangement of instrument specification which compatible to the indicator and good multiple-choice rules of writing. The development of the cognitive test instrument of this stage is referred to as the initial product.

#### 3.1.2 Design validation

Stages of design validation are carried out to determine the extent to which the level of product eligibility of the knowledge test is developed based on the data obtained by the evaluation expert. The results obtained from three experts are analyzed and then used as considerations in the form of prototype 1. The average design validation results by the evaluators consist of these aspects: substance, construction, language, objectivity and systematics can be seen in the following table [10].

### Table 2. Average scores of the test instrument content validation of basic concepts of science

| No | Substance   | Average of the Assessment | Percentage | Criteria       |
|----|-------------|---------------------------|------------|----------------|
| 1  | Substance   | 3.56                      | 89.0 %     | Very Eligible  |
| 2  | Construction| 3.67                      | 91.8 %     | Very Eligible  |
| 3  | Language    | 3.82                      | 95.6 %     | Very Eligible  |
| 4  | Objectivity | 3.33                      | 83.3 %     | Very Eligible  |
| 5  | Systematics | 3.67                      | 91.8 %     | Very Eligible  |

Table 2 shows that qualitative analysis of all the aspect such as substance, construction, language, objectivity and systematic aspects has a very good eligibility score in accordance with the specific criteria.
3.1.3 Revision Design

Revision is made based on the validation results analysis of designs as well as critics, suggestions and expert input evaluation. This revised result is called as prototype 2. Then, the revisions made to each of these aspects can be seen in Table 3.

Table 3. Critics, suggestions and expert input evaluation on the stage of revision design

| No | Critics, Suggestion and Expert Input | Revision |
|----|-------------------------------------|----------|
| 1  | Substance aspect                    |          |
|    | ▪ It is necessary to check and confirm QS Al Anbiya verse number 180 in question number 15 |          |
|    | ▪ Instrument seek to present questions that HOTS-oriented |          |
|    | ▪ The concept of integrated science and Islamic must be reproduced |          |
|    | ▪ Align the conformity of QS Al Anbiya verse 180 with the material content about in the question number 15 |          |
|    | ▪ The questions must adapt HOTS-oriented questions |          |
|    | ▪ The questions related to integrated science and Islamic study will be added gradually |          |
| 2  | Construction aspect                 |          |
|    | ▪ it is necessary to check writing error and fix it, construction of the question should be fixed in several number. |          |
|    | ▪ It is necessary to be consistent in the order and the term. |          |
|    | ▪ The unclear distractor homogeneity must be cleared. |          |
|    | ▪ It is necessary to add graphic/picture/illustration that suitable to the question. |          |
|    | ▪ The writing should contain subject, verb, object and adverbial as rules. |          |
|    | ▪ Careful of writing procedure and the use of the term. |          |
|    | ▪ Fixing the questions related to the writing procedure. |          |
|    | ▪ Fixing the questions instrument that focus on the writing procedure and the question construction. |          |
|    | ▪ Adjusting the order and the terms |          |
|    | ▪ Fixing distractor function |          |
|    | ▪ Adding graphic/picture/illustration in several questions. |          |
|    | ▪ Making the questions are adjusted with the applicable rules. |          |
|    | ▪ Paying attention and fixing the writing procedure. |          |
|    | ▪ Improving grammatical based on the writing rules. |          |
|    | ▪ Aligning question indicator. |          |
| 3  | Language aspect                     |          |
|    | ▪ Questions that are not in accordance with standard language rules should be corrected |          |
|    | ▪ Make clear the question instruction according to the question stage. |          |
|    | ▪ Fixing the questions using standard writing rules. |          |
|    | ▪ Fixing the instruction appropriately. |          |
| 4  | Objectivity aspect                  |          |
|    | ▪ More checking all the questions according to the function of the instrument |          |
|    | ▪ Paying attention to the question instrument that match with the indicators. |          |
| 5  | Systematic aspect                   |          |
|    | ▪ Adding the stage interpretation assessment result |          |
|    | ▪ Adding the interpretation result. |          |

3.1.4 Product Trial

Stages of the product trial begins with testing legibility to lecturers of Basic Concept of Science subject in Madrasah Ibtidaiyah Teacher-Training Program, simulating the use of cognitive learning outcomes and limited group testing.
3.1.4.1 Legibility Test
The legibility test instrument is done to the lecturer of Basic Concept of Science subject which is the one of research sample. The results of the legibility test given to the lecturer are shown in Table 4. shows the average result of the lecturers' legibility test is excellent so that it can be continued for simulation to the students.

Table 4. The Results of Legibility Test Instrument by Lecturer

| No | Indicator                                                                 | Average | Criteria  |
|----|---------------------------------------------------------------------------|---------|-----------|
| 1  | The instructions on the question and answering the questions are given in a coherent and clear manner | 5.00    | Very good |
| 2  | Conformity of the item with the achievement competency                   | 5.00    | Very good |
| 3  | The use of language in accordance with the writing standard rules and correct spelling | 4.75    | Very good |
| 4  | The clarity of the graphic/picture/illustration on the questions.        | 4.75    | Very good |
| 5  | The correct and valid answer keys                                         | 5.00    | Very good |

Average Score 4.92

3.1.4.2 Product Trial Simulation
The simulation of this product trial is conducted to students of Madrasah Ibtidaiyah Teacher-Training program class 3A. The students’ mastery of Basic Concept of science subject can be differentiate into three classes, those are student with highest ability, medium ability and lowest ability. The simulation results of product testing to students can be seen in Table 5. It shows that experimental simulation results of products instrument test on 15 students with different abilities can show actual capabilities from the students’ correct answer.

Table 5. results of product testing to students

| No | Class groups according to ability | Correct answers |
|----|-----------------------------------|-----------------|
| 1  | High ability                      | 28              |
| 2  | Medium ability                    | 22              |
| 3  | Low ability                       | 15              |

3.1.4.3 Product Trial
Product test is conducted to know the quantitative analysis of the problem. This test was conducted on the students of Madrasah Ibtidaiyah Teachers-Training Program which became the research sample. Quantitative analysis was performed using statistical analysis. Reliability score were carried out on 40 multiple choice questions using formula analysis [11]. Range gain with this formula is 0.423. Standard value of reliability is 0.43 it includes still in the range between 0.40< r ≤ 0.60 on some criteria. This condition is influenced by the scattering of group scores of the students being measured. The lower the distribution of questions, the lower the estimated coefficient of a question reliability. The difficulty level also gives support to the reliability score of the question. The more difficult question is the lower reliability one [12].

3.2 Cognitive Instrument Development Products in the Basic Concepts of Science
The resulting product from this development study is multiple choice questions with five possible answers as many as 40 items consisting of 14 standard indicators of competence listed in the lattice grid. Distribution of cognitive test instruments according to taxonomy Bloom, the revisions are 18 pieces about C4 question, 22 pieces about C5 questions, and 4 pieces about C6 questions.
3.2.1 The results of Cognitive Test Instrument Analysis in the Basic Concept of Science subject

Based on quantitative analysis of 40 tested items to the students of the third grade Madrasah Ibtidaiyah Teachers-Training Program obtained validity, reliability, discrimination power and difficulty level that shows in the Table 6.

Table 6. The result of cognitive aspect analysis of multiple-choice instrument on integrated science

| Criteria                                      | Number of multiple-choice questions | Note   |
|-----------------------------------------------|-------------------------------------|--------|
| Valid multiple-choice question                | 31                                  | Used   |
| Invalid multiple-choice question              | 9                                   | Not used |
| Question reliability                          | 0.434                               | Medium |
| Easy question                                 | 6                                   | Not used |
| Moderate question                             | 27                                  | Used   |
| Difficult question                            | 7                                   | Used   |
| Question with bad discrimination power        | 7                                   | Not used |
| Question with moderate discrimination power   | 8                                   | Used   |
| Question with good discrimination power       | 22                                  | Used   |
| Question with very good discrimination power  | 3                                   | Used   |

An analysis of the integrated learning outcomes resulted in 31 valid category questions and 9 invalid category questions used in the test. Integrated science learning outcomes are used if they fall into the four categories of analysis, both validity, reliability, discrimination power and difficulty levels. Analysis of 31 questions declared as valid questions with score $t_{count} > 1.70$, while 9 declared as invalid questions with score of $t_{count} < 1.70$. The condition of valid questions and invalid questions is expressed by comparing score of $t_{count}$ with $t_{table}$ [13].

The instrument of integrated science and religion learning assessment which developed has been feasible because it fulfills the criteria for content validity according to material experts, evaluation experts and science learning practitioners’ experts. Empirical testing of this instrument can be used in learning and produces good reliability. The practicality of the assessment instruments developed using questionnaires shows that all aspects in the category are very good. Thus, the instrument of the integrated science learning assessment practically can be used as an assessment tool. The lecturers of Sciences Lecture at Islamic college can apply the final product of an instrument of integrated science learning assessment to measure students’ competency on the basic concepts of science lecturing.

4. Conclusions and suggestions

The stages of the conducted development tests are product design, design validation by three experts which have validity score is very eligible, and the test product is very good for lecturers and students. There are 9 items that are invalid according to the quantitative analysis of the 40 items tested are questions number 1.4, 9, 12, 17, 26, 30, 32, and 36. There are 4 items to be revised in terms of multiple-choice those are the item number 8, 20, 38, and 40. There are 31 items that have good quality to measure students' ability.

Acknowledgement

Our great thanks you are conveyed to the Supervisor, Science Education Program State University of Semarang. Great family of academic community of Madrasah Ibtidaiyah Teachers-Training Program UIN Walisongo Semarang which is very helpful until the completion of this research.
References

[1] Akib E and Ghafar M N A 2015 *Int. Educ. Stud.* **8** (4) 166
[2] Nurwardani P, Tutupoho R R, Mulyono E, Nugroho S W, Endrotomo, Arifin S, Djajanto L, Soetanto H, Mursid S P, Sugiharto L, Wastutiningsih S P, Utama M S, Arifin P, Ghufron A, Evawany and Priyautama F 2016 *Panduan Penyusunan Kurikulum Pendidikan Tinggi* (Jakarta: Kemenristekdikti)
[3] Mardapi D, Kumaidi and Kartowagiran B 2011 *J. Penelit. dan Eval. Pendidik.* **15** (2) 326
[4] Scheiblechner H H 2009 *Psychol. Sci. Q.* **51** (2) 181
[5] Sukor N S, Osman K and Soh T M T 2013 *Asian Soc Sci.* **9** (16) 126
[6] Sugiyono 2015 *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif dan R&D.* (Bandung: Alfabeta)
[7] Arifin Z 2016 *Evaluasi Pembelajaran* (Bandung: Remaja Rosdakarya)
[8] Kasilingam G, Ramalingam M and Chinnavan E *J. Young Pharm.* **6** (4) 27
[9] Creswell J W 2012 *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research 4th ed.* (Boston, USA: Pearson)
[10] Freitas S, Prieto G, Simões M R and Santana I 2015 *Arch. Clin. Neuropsychol* **30** (5) 435
[11] Ghazali N H M 2016 *Int. J. Eval. Res. Educ.* (IJERE) **5** (2) 148
[12] Taherdoost H 2016 *Int. J. Acad. Res. Manag.* (IJARM) **5** (3) 28
[13] Looney A, Cumming J, van Der Kleij F and Harris K 2017 *Assess. Educ.: Princ. Policy Pract.* **25** 442