Increasing Teacher Professional Competence in Developing Procedural Abilities Using the Application of Assessment of Integrated Science Using Mobile Learning on Android Platform Gadgets

Allesius Maryanto[1], Dadan Rosana[1], Didik Setyawarno[1]

[1]Science Education Study Program, Faculty of Mathematics and Science, Yogyakarta State University, Jl Kolombo No 1, Karangmalang, Depok, Sleman, Yogyakarta, Indonesia

Corresponding Author. Email: allesius_maryanto@uny.ac.id

ABSTRACT

The development of smart mobile phone terminals, mobile learning has become an effective and efficient way for teachers to learn, so that the continuous development of teacher competency development can take place in a fun way. From the aspect of integrated science learning material, it is very clear that science learning in the 2013 curriculum is carried out on an integrated basis. Integrative science has the meaning of combining various aspects, namely the domains of attitudes, knowledge and skills. Therefore, the assessment system developed in science learning must also be able to develop four dimensions of knowledge (factual, conceptual, procedural). In this study, procedural knowledge was specifically developed using the assessment of integrated science application using mobile learning on gadgets with the android platform. The methodology developed in this study is a Research & Development spiral model as referenced by Cennamo and Kalk (2005: 6). In this spiral model, there are 5 (five) development phases, namely: (1) definition, (2) design, (3) demonstration, (4) develop, and (5) presentation (deliver). The research results that have been achieved are; (1) A model for enhancing the professional competence of teachers in the development of procedural ability tests through the design of an assessment of integrated science using mobile learning on gadgets with a proper android platform based on expert judgemnet, (2) Practicality tests of the model have been carried out based on the results of training involving junior high school teachers in the MGMP area of science teachers, Mlati District, Sleman Regency, (3) The model is effective in improving students' procedural abilities.

INTRODUCTION

The development of cell phone technology is so fast, and it has become a life style for most people, including teachers in Indonesia. Almost all teachers have cell phones, some of them feel that it is not enough just to have a cell phone, and many have also used an Android platform plug. This development, of course, must be utilized in the world of education which is then called mobile learning. Mobile learning is part of e-learning (Tsvetozar, Evgenia and Smrikarov, 2014; 34). Related to the large number of teachers using mobile devices, mobile learning can be used as an alternative to solving problems in the education sector, and one of them that will be developed through this research is in the field of learning evaluation in the form of text or images accompanied by sample questions. as well as interactive instruction to improve the quality of teachers to be better at making or delivering learning materials and managing teaching and learning activities.

Mobile learning in this case was developed to improve teacher professional competence, especially with the demands of implementing a new curriculum, Curriculum 2013. Mobile learning has several advantages including independence in

learning. But on the other hand, mobile learning tools have limited resources and a variety of platforms so that a design is needed that can guarantee compatibility and interoperability.

Starting in 2016, it is certain that the implementation of the 2013 curriculum, which has been evaluated and revised, will be implemented gradually throughout Indonesia. As one of the facilitators involved in the evaluation and revision of the 2013 Curriculum at the central level, both at the Center for Curriculum and Books, and at the technical level of the implementation of national facilitator training at the Directorate of Junior Secondary School Development, Ministry of Education and Culture, researchers have observed that one of the important aspects in Science learning is related to the application of a student-centered learning approach and its evaluation instruments. One of the biggest obstacles faced by teachers is the lack of ability to develop assessments to evaluate the dimensions of high-level knowledge which in the revised Bloom taxonomy are related to higher order thinking skills (CELT, 2011: 63).

As the academic community of an educational staff education institution (LPTK), researchers with the expertise of Educational Measurement and Evaluation feel challenged to contribute to preparing prospective professional teachers, one of the main competencies of which is to master the knowledge or subject matter they teach. Mastery of knowledge includes mastery of facts, concepts, procedures (Clark, D., 2010: 34). It is very important for a prospective teacher to demonstrate mastery of knowledge at least to the level of being able to explain. The level is higher than that if the teacher is able to predict the impact of the treatment of each action on the improvement of students' attitudes, skills and knowledge (Anderson & Krathwohl, 2001: 75). And the peak of mastery of knowledge is if the teacher is able to control his every action so that he knows its effect on students. The roadmap for the capacity and capability of prospective teachers is not much different from the roadmap for the capacity and capability of teachers as depicted in Figure 1.

Integrated science learning, oriented to produce prospective teachers who are required to have comprehensive professional abilities by combining science, physics, chemistry and biology, so that in accordance with policies in the implementation of the 2013 Curriculum, the main priorities of the dimensions of knowledge developed include procedural abilities and metacognitive abilities. Therefore, this study aims to improve the academic competence of science teacher candidates in developing an integrated assessment of the dimensions of procedural knowledge in science learning. This is very important related to the learning outcome in IQF level 7 (teacher professional program), namely mastering the theoretical concepts of certain fields of knowledge in general, and being able to formulate procedural problem solving, so that students are accustomed to developing high-order thinking skills.

The problems faced and experienced by teachers in developing dimensions of procedural knowledge are still not good, for example in practicum activities. Research results from the National Education Ministry's Balitbang (Rustad et al., 2004; Wiyanto, 2005) suggest that teachers' ability to design practicum is still low. Approximately 51% of junior high school science teachers in Indonesia cannot use the laboratory equipment available at their schools. Almost all problems in laboratory activities are part of the dimensions of procedural knowledge. An example of developing procedural knowledge in practicum activities can be seen in Figure 2 below! Meanwhile, metacognitive is a process of arousing interest because a person uses cognitive processes to reflect on their own cognitive processes (Kuntjojo, 2009). Metacognitive is very important because knowledge about cognitive processes can guide students in developing and choosing strategies to improve positive performance.

There are many factors that can cause problems in mastering the above dimensions of procedural knowledge. The teacher factor as a concept planter is very influential in this case. Lack of mastery of material and monotonous teaching methods can reduce the effectiveness of the conceptual process in students. Lack of affirmation of the perception of the concept or prerequisite material needed in learning can also lead to misconceptions in students' understanding, especially on sustainable material. In addition, the teacher's lack of knowledge on the level of student knowledge can also be an obstacle to the process of planting concepts. Furthermore, to get procedural knowledge that is well connected to metacognitive knowledge, of course, it is necessary to embed concepts accompanied by their relation to other concepts that have been studied and also requires practice to represent abstract ideas or concepts with symbols (Cracolice, etc., 2008), and exercises to use rules or procedures to solve problems in the science field based on the underlying concepts.

Based on this background, the objectives of the study are: 1) to produce a mapping of the academic competence of prospective teachers, especially the ability to develop assessment instruments using mobile learning technology with the Android platform for the dimensions of procedural knowledge in integrated science learning, 2) to obtain validity and reliability test results. items for measuring the dimensions of procedural knowledge in integrated science learning made by teachers and participants of the Science Teacher

Professional Program, and 3) Developing a model for increasing teacher competence in developing standardized assessments by utilizing mobile learning technology with an Android platform for measuring dimensions of procedural knowledge in integrated science learning.

METHOD

Educators must have academic qualifications and competencies as learning agents, be physically and mentally healthy, and have the ability to realize the goals of national education. The academic qualification referred to above is the minimum level of education that must be met by an educator as evidenced by a diploma and / or certificate of relevant expertise in accordance with the provisions of the applicable laws. Competencies as a learning agent at the primary and secondary education and early childhood education levels include; (1) Pedagogic competence, (2) Personality competence, (3) Professional competence, and (4) Social competence, as stated in the Regulation of the Minister of National Education of the Republic of Indonesia No. 16 of 2007 concerning Academic Qualification Standards and Teacher Competencies.

The assessment of learning outcomes by educators is carried out continuously to monitor the process, progress, and improvement of results in the form of daily tests, midterm exams, final semester exams, and class promotion exams. Assessment of learning outcomes by educators is used to assess the achievement of student competencies; materials for preparing reports on learning outcomes; and improve the learning process. This is regulated in Permendikbud number 104 of 2014. A fundamental difference with the previous permendikbud-permendikbud which contains assessment guidelines, namely Permendikbud number 66 of 2013 concerning assessment standards, Permendikbud number 81A of 2013 concerning implementation of the 2013 curriculum and most recently Permendikbud no 59 years 2014 concerning the 2013 SMA / MA curriculum which was passed on July 2, 2014, is on the range of values and writing numbers on report cards.

Referring to the research design, the research procedure is carried out through stages in research and development (R & D). The main objective of R & D is to develop and validate a program or model that will be used so that the objectives are effective and ready to be implemented. The stages of R&D are formulated into a 4-D model (Thiagarajan, 1975: 5) and adjusted by Borg and Gall (1983: 775). The problem that then arises in connection with this is the still weak ability of prospective teachers to develop an assessment system in accordance with the demands of the curriculum. Because it is

necessary to improve the academic competence of prospective teachers in the development of assessment of the dimensions of procedural knowledge in integrated science learning, with the research design shown in the following stages.

a. Define (D-1) / Research and Information Collection

1) Theory Analysis / Literature Study

This stage analyzes the theoretical model model for equalizing test quality, including:

☐ There are three types of equalization test designs that can be used, namely the single group design, the equivalent group design, and the anchor test item design. In a single group design, one group of participants who responds to two sets of tests (X and Y) is used. Item parameters from the two test kits were estimated separately by calibrating the test taker's ability parameters or item parameters. Based on this design, by calibrating the participant's ability parameters, the item parameters of the X and Y test kits are already on the same scale.

☐ Ideal to equalize the scores of several test sets, then the test-set test sets are given to the same respondent. By comparing the abilities of the test takers from two / more test sets, an equalization of the two test sets can be done. In reality in the field, this design is difficult to do due to fatigue, learning, and practice factors for the second or next test. In addition, there will be difficulties in planning enough time for respondents to take the test more than once (Miyatun and Mardapi, 2000).

☐ In the equivalent group design, two groups of equivalent participants (K1 and K2) and two test sets (X and Y) were used. The group of participants K1 did the test set X and the group of participants K2 worked on the test set Y. Given that groups K1 and K2 are equivalent, the two groups are considered to be single. Subsequent determination of the conversion constant is like a single group design. The advantage of this design is that it avoids the negative effects caused by training and fatigue of test takers, while the disadvantage is that there is a possibility of bias due to the fact that it is not easy to make the distribution of the ability of the two groups of test takers completely equivalent (Sukirno, 2007: 310).

☐ Anchor test design is usually used when the safety issue of the test is an important consideration and it is possible to carry out several tests at one time. In this design, each test set has several common items and each group works on a different test set. In this design, there are two variations, namely (Chong and Osborn, 2005): 1) if the common item is taken into account in the scoring it is called the internal common item; and 2) if the common item is not taken into account in the scoring it is called the external common item.

2) Task Analysis / Needs Assessment
This stage identifies the main process skills and analyzes them in the required sub-skill sets. This analysis ensures the comprehensiveness of the tasks in determining the passing of the invitation line Teacher Competency Test system. Research activities include the development of a test design, containing the objectives of the assessment to be carried out, the time (time taken) for the implementation of the test, the main message of the curriculum (learning objectives and the outline of the topic of the test material), item indicators (characteristics of test material mastery and achievement learning objectives), as well as the number and shape of the items (per indicator, per topic, and the whole test). The distribution of the items in the test should pay attention to the balance of mastery demands according to the message of the curriculum, so as to provide a nuance of representation of the topics discussed.

3) Concept Analysis / Needs Assessment
This stage identifies the main concepts being taught, arranges them in a hierarchy and describes them in the main themes. The test construction steps taken are as follows: 1) setting the test objectives; 2) curriculum analysis; 3) analysis of textbooks and other learning material sources; 4) making grids; 5) writing specific instructional goals; 6) writing questions; 7) study questions (face validity); 8) limited test reproduction; 9) test trials; 10) analysis of trial results; 11) revising questions, and l2) assembling questions into tests. The equivalent method according to grain response theory serves to determine the conversion constant. This is because the equalization between two or more test sets can be done if the conversion constant is known. The resulting conversion value is then substituted in the scale equation in the equivalent design used. There are several test equalization methods that can be used and factors influence the accuracy of the test equalization method. In item response theory, there are four test equalization methods, namely: regression, mean sigma, mean and sigma rigidity, and characteristic curves.

b. Design (D-2)
The first test equalization method is the regression method. Determination of the conversion constants a and b using the regression method is carried out by taking into account the responses of the test takers on the two test sets X and Y. Estimates of item parameters and participant ability parameters meet the linear regression equation as follows:

$$y = ax + b + e \text{ with } a = r_{xy}\, S_y/S_x \text{ and } b = \hat{y} - ax$$

Explanation:

Y : estimated ability or item parameter estimation on the Y test device
X : estimated ability or item parameter estimation on test device X
r_{xy} : the correlation coefficient between X and Y
y, x : mean of y and x
Sy, Sx : standard deviation of x and y
e : error in the regression error estimate

The use of this method is non-reciprocal (asymmetric) so it is not sufficient for determining the conversion constant, especially considering that the equalization of two or more test sets requires invariance and reciprocity conditions from the equivalent test set.

c. Develop (D-3) / Develop Preliminary form of Product
This stage obtains approval to improve the quality of the assessment used in determining the passing of the invitation-only Teacher Competency Test. A number of experts were asked to evaluate the IRT that had been designed, including four equalization test methods, namely: regression, mean sigma, mean and rigid sigma, and characteristic curves then based on expert feedback, the model that had been produced was modified / revised to be more precise, effective, and useful and high quality technique.

d. Dessiminate (D-4)
Testing of development results devices (Preliminary Field Testing) is carried out in collaboration with the Department of Education and Culture and State Universities through the trial phase and has been revised (Main Product Revision), then the tools are applied in making the actual procedural ability test (Main Field Testing), then observed all the variables that become the focus / development goals.
In the initial field trial steps and limited trials the model of increasing the academic competence of teachers in the development of assessment of the dimensions of procedural knowledge in integrated science learning, the research subjects were used as junior high school science teachers in Mlati sub-district, Sleman Regency, Yogyakarta.

RESULTS AND DISCUSSION
Current understanding of learning focuses on active, cognitive and constructive processes that are incorporated into meaningful learning. Students in this case act as individuals who are active in each of their learning; they can select information constructed by their own understanding of the selected information. Science teachers are not passive recipients, recording information obtained from their environment, lecturers, textbooks or computer and android based information media.

This is a change from a passive view of cognitive learning and a constructive perspective that emphasizes how science teachers know (knowledge) and how they think (cognitive processes) about what they know as long as science teachers do meaningful learning.

Given the many types of knowledge, especially in the development of cognitive psychology, it can generally be classified into four types of general knowledge, namely Factual, Procedural, Metacognitive,. However, this study is specifically limited to science teacher candidates on procedural and metacognitive abilities.

The dimensions of knowledge developed in this study and examples are shown in Table 1. (Anderson, et.al., 2001).

Table 1. Types, Subtypes, and Examples of Dimensions of Procedural Knowledge

Procedural Knowledge:
Studying the relationships between elements in a large structure that allows the elements to function together.

Sub Type	Examples developed in research (Question Number)
1. Knowledge of classification and categories	Classification and categories of force types and examples of simple planes according to the workings of bones and joints (5,7, 19, 20, 21, 23, 24, 25, 26,42,)
2. Knowledge of principles and generalizations	Principles and generalizations of the concept of fluid pressure in human organ systems (8,6, 28, 31, 32, 33, 34, 35,)
3. Knowledge of theories, models, and structures	Knowledge of the application of the theory of mikanika in the motion system of living things (11, 13, 14, 15, 16,18, 22)

Sub Type	Examples developed in research (Question Number)
1. Knowledge of skills in a specific area and algorithms	Problem solving algorithms in the system of equilibrium of objects and force vectors in the body (1,3, 12, 37, 38, 39, 40, 41, 43)
2. Knowledge of techniques and methods in a particular field	The right method for solving optical problems and waves in the eye and hearing senses (2,4,17, 44, 45, 46, 49,50)
3. Knowledge of the criteria for determining when to use appropriate procedures	Criteria for determining when must apply Legal procedures Newton in living systems (27, 29, 30, 36, 47, 48)

The knowledge of classifications and categories developed includes the categories, divisions and arrangements used in different materials. This knowledge generally reflects how the experts think and solve their problems, where special knowledge is important for the problems that have been solved. Knowledge is an important aspect in developing an academic discipline. Examples of developed knowledge are; (1) knowledge of the various types of forces that act on living things, (2) knowledge of various forms of energy transformation, (3) knowledge of the parts of the body organs (excretory system, movement system, respiratory system, etc.) , and (4) knowledge of different kinds of environmental problems.

The basic and general knowledge developed includes real abstractions that conclude the research phenomenon. This abstraction has enormous value in describing, predicting, explaining or determining the most appropriate and relevant action or direction to be taken (Cracolice, at.all., 2008). Examples of developed knowledge are; (1) knowledge of major generalizations about climate change, (2) knowledge of basic science laws, (3) knowledge of the basics of chemistry that are relevant in biological processes, and (4) knowledge of the basic principles of applying science in biological systems.

Knowledge theories, models and structures that are developed include basic knowledge and generalizations with clear interrelationships, a systematic view of a complex phenomenon, problem, or material. This knowledge is an abstract formula. Examples of developed knowledge are; (1) knowledge of the reciprocal relationship between science principles as a basis for developing science theory, (2) knowledge of organ structure as a whole (organization, function), (3) knowledge of evolution, (4) knowledge of plate tectonic theory, and (5) knowledge genetic model (DNA).

Metacognitive knowledge is knowledge of how to do something. Such as knowledge of skills, algorithms, techniques, and methods which are collectively known as procedures. Increasing the ability of prospective teachers through laboratory activities can increase their generic abilities (Rosnita, 2016). Laboratory activities are very dominant in developing metacognitive knowledge.

a. General-Specific Skills Knowledge and Scientific Work

Scientific work is used with various learning activities both in the classroom, laboratory and in the environment or nature. The scientific approach procedure, when applied, generally results in answers to hypotheses related to natural phenomena. Although this is done in metacognitive knowledge, the result of this metacognitive knowledge often becomes factual or procedural knowledge (Clark, 2010).

Scientific work (eg through practicum) to get answers to questions about the relationship between force, mass and acceleration is metacognitive knowledge, the answer is Newton's second law (F = ma) as easy as factual knowledge. Again, the emphasis here is based on the student's understanding of understanding and solving it himself, but on the procedures for how he or she derives these equations. Examples of developed knowledge are; (1) knowledge of skills in designing Newton II law experiments, (2) knowledge of skills used in understanding how muscles work based on force analysis, and (3) skills knowledge of various strategies to solve environmental problems

b. Knowledge of Special Methods and Techniques

Knowledge of specific methods and techniques includes a very broad knowledge of the results of consensus, agreement, or disciplinary norms rather than knowledge that is more directly the result of observation, experiment, or discovery. Examples of developed knowledge are; (1) knowledge of relevant research methods for science, (2) knowledge of techniques used by scientists in finding solutions to problems, (3) knowledge of methods for evaluating

environmental health concepts, (4) knowledge of various waste treatment methods .

c. Knowledge of the Criteria to Determine the Appropriate Use of Procedures

Examples of developed knowledge are; (1) knowledge of the criteria for determining several types of bone in the skeletal system of the human body, (2) knowledge of the criteria for determining the methods used in solving the Archimidhes law equations, (3) knowledge of the criteria for determining statistical procedures for using data collected in experiments, and (4) knowledge of criteria to determine the techniques for implementing and constructing a waste treatment system.

In this study, to equip science teachers for prospective teachers who carry out PPL, a model is developed for independent learning on how to develop procedural proficiency tests. The questions analyzed were 40 items from the 50 questions made. 10 questions are not used because in terms of the dimension of knowledge it leads to factual abilities. The empirical analysis of the characteristics of the tests and items based on the classical approach includes the difficulty level of the items, the difference between the items, the reliability of the tests and the measurement errors. To simplify calculations, testing using Cronbach Alpha was used to test the reliability of the instrument at the time of the trial (Rosana, et al., 2017). Table 9 describes in detail the conditions of each multiple choice item in terms of difficulty level, difference power, distractor state that needs to be revised or not, and explains the decision on each item. An item is said to be "good" if it meets three conditions: (1) moderate difficulty level (0.3 ≤ difficulty level ≤ 0.7); (2) the discrepancy of items that can distinguish (Distinctive Power> 0.3); and (3) at least the three cheaters were selected by 2% of test takers. The results of the difficulty level analysis show that; (1) 19 items (38%) that were not good based on the level of difficulty, namely items 1, 2, 7, 8, 12, 14, 17, 18, 22, 23, 27, 29, 30, 32, 35 , 42, 44, 47,48, (2) 12 items (24%) that could not distinguish the ability of science teachers, namely items 1, 2, 17, 18, 22, 27, 29, 30, 37, 39 , 47 and 48, and (3) 4 items (8%) whose response distribution did not work, namely items 9, 18, 29, and 48.

The results of the item characteristics analysis showed that the difficulty level of the items on the Natural Science Procedural Ability Test instrument started from 0.193 to 0.939 with a mean of 0.613. The average item difficulty level is in a good category because according to Crocker & Algina (1986: 313), Ridho (2007) and Wright & Linacre (2008: 227) for multiple choice with five alternative answers, the optimal level of difficulty for the items is 0.59 . Furthermore, referring to Allen & Yen

(1989) criteria, the difficulty level categorization on   each item is as follows.

Table 2. Categorization of Item Difficulty Level

Category	Percentage (%)	Item Number
Easy	20 (10 items)	1, 2, 7, 8, 18, 22, 27, 29, 30,48
Moderate	72 (36 items)	3, 4, 5, 6, 9, 10, 11, 13, 14, 15, 16, 19, 20, 21, 23, 24, 25, 26, 28, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 44, 45, 46, 49,50
Hard	8 (4 items)	12, 17, 42, 47

The difference power of the IPA Procedural Ability Test equipment starts from 0.156 to 0.581 with a mean of 0.383. The item difference power of the analysis using the classical approach (biserial point correlation) shows that there are 10 items (25%) that are not able to distinguish the ability of science teachers. This is because the ten items have a different power index below the criteria referred to, namely 0.3 (Reynolds, Livingston, & Wilson, 2010; Kartowagiran, 2011).

The data obtained in this study were the mastery of material in the initial and final Procedural Ability Tests for science teachers on junior high school science material obtained from the pretest and posttest results. The pretest was given to the science teacher before learning the development of the junior high school science procedural ability test was given to determine the science teacher's mastery of the initial material. Posttests are given to science teachers after the learning process for developing junior high school science procedural ability tests is completed to determine the science teacher's mastery of material after receiving treatment. Below is a description of the results for each data.

Initial data on the procedural ability of science teachers in junior high school science material can be seen through pretest. The pretest is in the form of a test of cognitive and psychomotor learning outcomes for science teachers. There were 50 pretest questions given to the experimental group and the control group. In summary, the data on the initial ability of science teachers can be seen in Table 3.

Table 3. Data Parameters of the Science Teacher Procedural Ability Pretest

Variable	Nilai			
	Max	Min	Average	Dev. Standar
Control Class	52	30	42	5,29
Experiment Class	60	30	45	6,35

The preliminary data on the Procedural Ability Test for Science teachers on SMP science material is data obtained from posttests. Posttest was carried out in the control group and the experimental group. The questions used for posttest are the same as for pretest questions, the difference is the randomized number of questions. In summary, the data can be presented in Table 4.

Table 4. Postes Data Parameters for the Procedural Ability of Science Teachers

Variable	Nilai			
	Max	Min	Average	Dev. Standar
Control Class	86	58	69	6,82
Experiment Class	92	64	79	8,53

Hypothesis testing is done using the Manova test. From the above analysis, the data is known to be normally distributed and homogeneous and independent. This manova analysis used the assistance of the SPSS 21.0 program.

Hypothesis testing is carried out on procedural ability data and metacognitive abilities. Based on the results of the calculations, it can be seen that the F test is significant at α 5% so that it is not equal to 0. From the table it can be seen that the procedural ability is influenced by the availability of models for independent learning about the development of procedural ability tests. After it is known that the multivariate test is significant, the next step is the univariate F test.

Table 5. Levene's test of equality of error variance

Test Type	F	df1	df2	Sig.
Procedural Ability Tests	3.516	1	56	.066

From this table it can be seen that the significance value between procedural abilities is not the same. Namely the significance of Procedural ability of 0.066. In addition, the manova test can be seen in the following table.

Table 6. Test of between-subjects effect

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Procedural Ability Tests	.093[a]	1	.093	5.158	.025	.083
Intercept	Procedural Ability Tests	15.253	1	15.253	833.988	.000	.936
Treatment	Procedural Ability Tests	.093	1	.093	5.158	.026	.083
Error	Procedural Ability Tests	1.024	56	.019			
Total	Procedural Ability Tests	16.371	58				
Corrected Total	Procedural Ability Tests	1.118	57				

Manova analysis lies in whether the independent variable affects the dependent variable. This can be seen from the corrected model and treatment. Judging from the table, both of them produce the same F test information because both are tests of whether the independent variable affects the dependent variable. The results of the univariate F test on the treatment have a significance level of less than 0.05, thus the use of the model affects procedural abilities and metacognitive abilities. The Partial Eta Square (PES) values of procedural and metacognitive abilities were 0.093 and 0.746, respectively. This means that the model use model affects procedural abilities by 9.3% and metacognitive abilities by 74.6% (Holland, & Dorans, 2006).

From the results of this analysis, it can be seen that learning using a model is able to affect procedural abilities by 9.3% and learning models using learning models can affect metacognitive abilities by 74.6%. The results of this analysis can explain the involvement of science teachers to participate in learning by implementing learning using a model which is one indicator of learning effectiveness. Science teachers not only receive material from lecturers, but science teachers also try to explore and develop themselves. Learning outcomes not only produce value but can increase knowledge and concepts of science and metacognitive and procedural abilities.

The ability of science teachers to develop and work on multiple-choice questions with a high level of thinking needs to be trained, so that the application of learning using models can be optimal. For that we need not only hard-skills but also soft-skills for hard work and smart work in groups. This is in line with Rosana's (2014) research that science teacher soft-skills can be improved with context-based learning, for example the application of metacognitive knowledge in science learning.

This ability will be clearly seen if the science teacher is able to ask questions independently or in groups. The ability of science teachers to work on these questions can be detected through their ability to explain the completion of practice questions (Kim, at.all, .2008). In accordance with the principles of the learning model development model of procedural ability tests involving science teachers actively in the process of teaching and learning activities because this learning model requires science teachers to make their own questions and answers based on questions given by lecturers through stimuli in the form of pictures, stories or stories, diagrams, exposures and etc. Science teachers also try to explore and develop on their own, by applying learning using a model that encourages science teachers to try their best. The application of learning uses a model, science teachers are not only enthusiastic about doing science exercises so that learning outcomes not only produce science values and concepts, but teach them to have good discussions with their friends. Discussions require science teachers to convey ideas, ideas, opinions. Research by Purwoko, et al (2017) shows that the frequency of involvement of science teachers in learning activities increases teacher competence. In addition, science teachers also learn to respect other people's ideas, ideas, and opinions.

CONCLUSION

Based on the description of the research results and discussion described above, the following conclusions can be drawn; (1) the results of the univariate F test on treatment have a significance level of less than 0.05, thus the use of the model affects the procedural abilities and metacognitive abilities. The Partial Eta Square (PES) values of procedural and metacognitive abilities were 0.093 and 0.746, respectively. This means that the use of the model affects procedural abilities by 9.3% and metacognitive abilities by 74.6%, (2) the academic competence of prospective teachers, especially the ability to develop assessment instruments for the dimensions of procedural and metacognitive knowledge in integrated science learning, is not very satisfying because only 12 , 5% are able to get a test score above 60 (from a maximum score of 100) and increase to 23% after learning to use the model, (3) the validity and reliability of the items for measuring the dimensions of procedural and metacognitive knowledge in integrated science learning made by the teacher Science of prospective teachers is still low, so further research is needed to increase their competence, and (4) models for increasing the competence of prospective teachers in developing measurements of the dimensions of procedural and metacognitive knowledge, using models alone have not shown a significant improvement, still need to be given further treatment.

REFERENCES

Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing.* New York: Longman.

Allyn & Bacon (1995). Assimilation and Accomodation in Cognitif Development. http://www.abacon.com/slavin/ill.html diakses 16 Desember 2008

Aman, C., et.al. (2007). Student *Learning Teams: Viepoints of Team Members, Teachers and an Observer.* Vol 2 isuue| 2007, engineering education, pp. 1-12. Tersedia : garypoole@ubc.ca. [19 Maret 2008]

Asmin. 2014. Implementasi Teori Responsi Butir dan Fungsi Informasi Butir Tes dalam Pengujian Hasil Belajar Akhir di Sekolah. *Jurnal Pendidikan dan Kebudayaan,* X (48): 234-245.

Barnard. John. J. (1996). *In Search for Equity in Educational Measurement: Traditional Versus Modern Equating Methods.* Makalah: Disampaikan pada ASEESA National Conference di HSRC Conference Centre. Pretoria: Afrika Selatan.

Camilli, Gregory, dan Lorrie A. Shepard. (1994). *Methods for Identifying Biased Test Items.* California: Sage Publication.

Center for Excellence in Learning and Teaching (CELT). (2011). A model of learning objectives. Iowa State University. Retrieved March 2011, from http://www.celt.iastate.edu/teaching/RevisedBlooms1.html.

Clark, D. (2010). Bloom's taxonomy of learning domains: The three types of learning. *Big Dog & Little Dog's Performance Juxtaposition.* Edmonds, WA: Author. Retrieved from http://www.nwlink.com/~donclark/hrd/bloom.html.

Cracolice, M.S., Deming, J.C. & Ehlert, B. (2011). Concept learning versus problem solving: a cognitive difference. *Journal of Chemical Education.* 85 (6), 873-878.

Chong Ho Yu dan Sharon E. Osborn. (2015). Test Equating by Common Items and Common Subject: Concepts and Applications. *Practical Assessment, Research & Evaluation.* X (4): 187-198.

Crocker, Linda, & Algina, James. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart and Winston, Inc.

Dorans, N. J. (2004). Equating, concordance, and expectation. *Applied Psychological Measurement,* 28 (4),227-246.

Grinnel, Jr., Richard M. (1988). *Social Work Research and Education.* Thirt Edition.Canada: F.E. Peacock Publisher, Inc

Gronlund, Norman. E. 1985. *Measurement and Evaluation in Teaching.* New York: MaAssessment of Integrated Science using Mobile Learning pada Gadget Ber-platform Androidllan Publishing Company.

Hambleton, Ronald K, Swaminathan, H., dan Jane Rogers, H. 1991. *Fundamentals of Item Response Theory.* London: SagePublications.

Hambleton, Ronald K., dan Swaminathan, H. 1985. *Item Response Theory: Principle and Applications.* Boston: Kluwer Nijhoff Publishing.

Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.),*Journal of Educational measurement* (4th ed., pp. 187{220). Westport, CT: Greenwood.

Jihad, Asep, Abdul Haris. 2011. *Evaluasi Pembelajaran.* Multi Pressindo: Yogyakarta.

Kim, S., von Davier, A. A., & Haberman, S. (2008). Small-sample equating using a synthetic linking function. *Journal of Educational Measurement,* 45, 325{342}

Kolen, Michael J., dan Robert L. Brennan. 2004. *Test Equating, Scaling, and Linking: Methods and Practices.* New York: Springer.

Kolen, Michael J., dan Robert L. Brennan. 1995. *Test Equating.* New York: Springer Verlag.

Kumaidi. 2000. Standardisasi Butir Soal. *Jurnal Pendidikan dan Kebudayaan.* V(5): 132-143.

Livingstone, S. A., Doran, N. J. dan Wright, N. K. 1990. What Combination of Sampling and Equating Methods Work Best?. *Applied Measurement in Education.* III (2): 73-95.

Livingston, S. A., & Kim, S. (2009). The circle-arc method for equating in small samples. *Journal of Educational Measurement*, 46, 330{343}

Lord, F. M. (2009). The standard error of equipercentile equating. *Journal of Educational Statistics*,7, 165{174}

Lord, Frederick, M.1990. *Aplications of Item Response Theory to Practical Testing Problems.* New Jersey: LawrenceErlbaum Associates, Publishers.

Mary J.Allen and Wendy M Yen, 1989, *Introduction to Measurement Theory,* California: Broke.

McDonald, Roderick P. 1991. *Test Theory: A Unified Treatment.* New Jersey: Lawrence Erlbaum Associatiates Publisher.

Naga, Dali, S. 1992. *Pengantar Teori Sekor Pada Pengukuran Pendidikan*. Jakarta: Besbats.

Nitko, Anthony. J. 1992. *Criterion Reference Testing Workshop*: *Handouts and Reading Material Tidak dipublikasikan).* Cipayung, Bogor: Examination Development Unit (Puslitbang Sisjian).

Miyatun, Erna., dan Djemari Mardapi. 2000. Komparasi Metode Penyetaraan Tes Menurut Teori Responsi Butir. *Jurnal Penelitian dan Evaluasi.* II (3): 124-132.

Peraturan Pemerintah No. 19 Th 2005 Tentang: Standar Nasional Pendidikan (SNP). Bandung: Citra Umbara.

Peterson, N.S., Kolen, M.J., dan Hoover, H.D. 1989. Scaling, Norming, and Equating. In R.L. Linn (Ed), *Educational Measurement.* New York: MaAssessment of Integrated Science using Mobile Learning pada Gadget Ber-platform Androidllan.

Rahayu, Wardani. 2008. Pengaruh Metode Linking Terhadap Banyak Butir False Positive pada Pendeteksian DIF Berdasarkan Teori Responsi Butir. *Disertasi.* Jakarta: Universitas Negeri Jakarta.

Ridho, Ali. 2007. Karakteristik Psikometrik Tes Berdasarkan Pendekatan Teori Tes Klasik dan Teori Respon Aitem. *Jurnal Insan Media.* II (2): 1-28.

Setiadi, Hari. 1998. Bank Soal yang Dikalibrasi dengan Konsep IRT Memecahkan Permasalahan Ujian-ujian Sistematik yang Diadakan pada Periode-periode Tertentu, *Jurnal Kajian Dikbud* IV (13).

Setiadi, Hari. 2009. Permasalahan dan Solusinya dalam Pelaksanaan Ujian Nasional di Masa Mendatang, *Matahari: Jurnal Penelitian dan Pendidikan.* X (1): 66-74.

Skaggs, G. (2005). Accuracy of random groups equating with very small amples. *Journal of Educational Measurement*,42, 309{330}

Susongko, Purwo. 2005. Penyetaraan Parameter Butir Secara Konkuren untuk Menguji Secara Statistik Keberadaan Item Function (DIF). Makalah: Disampaikan pada Seminar Nasional *Hasil Penelitian tentang Evaluasi Hasil Belajar serta Pengelolaannya.* Pascasarjana UNY Didukung oleh Direktorat P2TK & KPT dan HEPI, Yogyakarta, 14-15 Mei 2005.

Sukirno, D. S. 2007. Penyetaraan Tes UAN: Mengapa dan Bagaimana. *Jurnal Cakrawala Pendidikan.* XXVI (3): 305-321.

Syarifah. 2007. Persyaratan Analisis Instrumen Sebagai Prasyarat Ketepatan Hasil Analisis Dalam Penelitian Pendidikan. *Cakrawala Pendidikan.* XXVI (2): 15-27.

Swediati, Nonny. 1997. *Metode untuk Penyetaraan* (*Equating*) *Sekor Tes Secara Klasik.* Pusat Pengujian Balitbang Dikbud: Jakarta.

Tumilisar, A.V.J. 2006. Akurasi Relatif Penyetaraan Sekor Tes untuk Sampel Berukuran 300 Ditinjau dari Metode Penyetaraan dan Teknik Penghalusan. *Jurnal Pendidikan Penabur.* V (6): 1-19.

Zhu, W. 1998. Test Equating: What, Why, How?. *Research Quarterly for Exercises and Sport.* Wayne State University.