

DEVELOPING COGNITIVE DIAGNOSTIC TESTS ON LEARNING OF SCIENCE

Yuli Prihatni
Sarjanawiyata Tamansiswa University, Yogyakarta
Email: yuliku7781@gmail.com

Abstract

A cognitive diagnostic test is required in learning activities to diagnose the ability and students' weaknesses through the stage of thinking. With diagnostic test is expected to the aim of learning that have not been known by the students can be identified and known the problems. The purpose of this study is to develop a diagnostic test of cognitive science and to obtain characteristics of cognitive diagnostic test in science subjects. The research used the development of diagnostic tests. The trials conducted in seven schools that have with high category, medium and low school category. Test subjects of the seventh grade students were 484. Development of data analysis based on descriptive cognitive diagnostic test done through Focus Group Discussion (FGD). Selection of test items is to obtain fit tests using the Quest program. Through FGD produces Learning Continuum, grating test, test specifications and item 28 in the form of multiple choices with reasons. By using the Quest program obtained 27 items fit is to see that the items are in a vertical line or have infit MNSQ between 0.77 until 1.30. Instrument reliability value is 0.74. The research instrument is compliant with the statistical Fit with an average value of infit MNSQ 1.01 with a standard deviation 0.09.

Keywords: *Diagnostic Tests, Cognitive, Science*

Introduction

The use of current diagnostic tests is very crucial because the teachers will be informed about the level of students' learning difficulties. Therefore, teachers can provide appropriate support and in accordance with what is required by the students when students have difficulty. Currently, students' failure is often considered normal and seen as normal. This can be seen by the many problems of education in Indonesia. One of these problems involve students is as follows the high rate of repeat and a lot of students who drop out of school, especially at the secondary school level, it still reflects the presence of a variety of issues related to inadequate quality of education.

According to Ridwan Efendi (2010) Indonesia has three times participated only for the grade eight students of junior high school (eighth grade students of SMP / MTs) in TIMSS (Trends of International on Mathematics and Science Study) in 1999, 2003, and 2007. The achievement of eighth grade students in Indonesia to three times participation in TIMSS is in the bottom than the achievement level of students in several countries in Asia (Hong Kong, Japan, Korea, Taiwan, Malaysia, Thailand). The average score Indonesian student

science achievement in TIMSS 1999, 2003, and 2007 respectively are 435, 420, and 433. With the scores of students Indonesia ranks 32 out of 38 countries (1999), ranked 37 of 46 countries (the 2003), and ranked 35 of 49 countries (2007). The average score on the TIMSS 2007 Indonesian students is lower from the average score of 500, and only reach the Low International Benchmark. With these achievements, the average Indonesian students are only able to recognize some basic facts but have not been able to communicate and combine the various topics of science, especially, applying the concepts of complex and abstract.

Science, according to Titus (1959: 78) contains three definitions are as numerous scientific disciplines, as a group of knowledge, and as methods. Besides, it is also confirmed that science is a series of related to concepts and developed from the results of experiments and observations. According to Hungerford, Volk & Ramsey (1986: 8) science is 1) the process of obtaining information through empirical methods; 2) information is obtained and arranged through the investigation logically and systematically, 3) a combination of critical thinking processes that produce reliable information and valid.

Diagnostic tests are needed in learning as well as in learning science. Diagnostic tests according to Ebel (1979: 375) are designed to determine the specific deficiencies or failure in learning in some subjects or lessons such as reading and arithmetic. A similar opinion was expressed by Gronlund states that the diagnostic test is a test designed to determine the cause of failure in teaching learners. Diagnostic tests have two main functions, namely: 1) In identifying the problems or difficulties experienced by students, 2) follow-up plan in the form of efforts to solving the corresponding problems or difficulties that have been identified.

Weeden, Winter & Bbroundfoot (2002: 20) states that the diagnostic tests are test to find out indication how far the difference between prior knowledge and skills expected is or tests is used to identify specific problems experienced by students. Embretson (2002: 221-222) suggests two reasons why it is necessary to develop cognitive psychology in the development of the tests. It is because: 1) The concept of construct validity gave little space on the cognitive theory in the development of the tests. 2) Development of tests is not accustomed to using a procedure involving cognitive theory. Mehrens & Lehmann (1973: 410) argues that the diagnostic tests are part of the achievement tests, diagnostic tests as well as from the obtained value is an achievement or student learning outcomes. However, as the main purpose of diagnostic tests do not see the students' achievement but to know the difficulties faced by students. The results of the diagnostic tests provide information on concepts that already exist or have not been understood. The concept is not understood well due to the occurrence of misconceptions one experienced by students. According to Mehrens

& Lehmann (1973: 410) in developing diagnostic tests must offer two assumptions: 1) The test is able to analyze the skills or knowledge to be included in the component skills or knowledge, and 2) Item test developed is able to measure the skills or knowledge. Diagnostic tests are used in the diagnostic assessments is good if it does not just reveal the students' understanding of the meaning of the lesson. The test can show and map out the weaknesses of students and knowledge that is not steady. Gorin (2007: 174) states that diagnostic tests should be able to uncover why students responded as they did. Diagnostic tests are more specific and more detailed. Ultimately, the diagnostic tests are tests that are used to determine student weaknesses. In developing diagnostic tests, the first step is to define the scope of the concept creating test items. Determination of scope in tune with the concept of the subject to be diagnosed in order to test items actually measure what it intends to measure.

A measurement model that is used in developing diagnostic tests on the learning of science is the Partial Credit Model (PCM). PCM was chosen to evaluate the data obtained through testing instrument diagnostic tests of cognitive science subjects in SMP. Partial Credit Model (PCM) is one of the Polytomus IRT models that is developed by Masters (1982) based on the model of Rasch dichotomous response of grain into models response polytomus. PCM as an extension of the Rasch model 1-PL models can use a sample that is not as big as when doing the calibration data is polytomus in model 2-PL or 3-PL (Keeves & Masters, 1999, 12-13).

The aim of this research is to develop a diagnostic test of cognitive science and to obtain characteristics of cognitive diagnostic test in science subjects.

Methods

This study uses diagnostic tests development research. Five-step development of diagnostic tests aimed at cognitive assessment according to Nichols (1994: 587) is 1) Based on the construction of the substantive theory. Substantive theory is the basis for the development of a test based on the research or review of research. 2) Selection of design. Design of measurement are used to construct items that can respond either by the person taking the test is based on knowledge, specific skills or other characteristics according to the theory. 3) Administration of the test. Administration of the test covers several aspects of the format item, the technology used to make the assay, the environmental situation at the time of testing, and so on. 4) Scoring results of the test is the determination of the value of tests which have been carried out. 5) Revision, the process of adjustment between theories and models, whether the tests were developed to support the theory or not.

The step of this research is from the beginning to the end of the activity. It can be seen in the following figure.

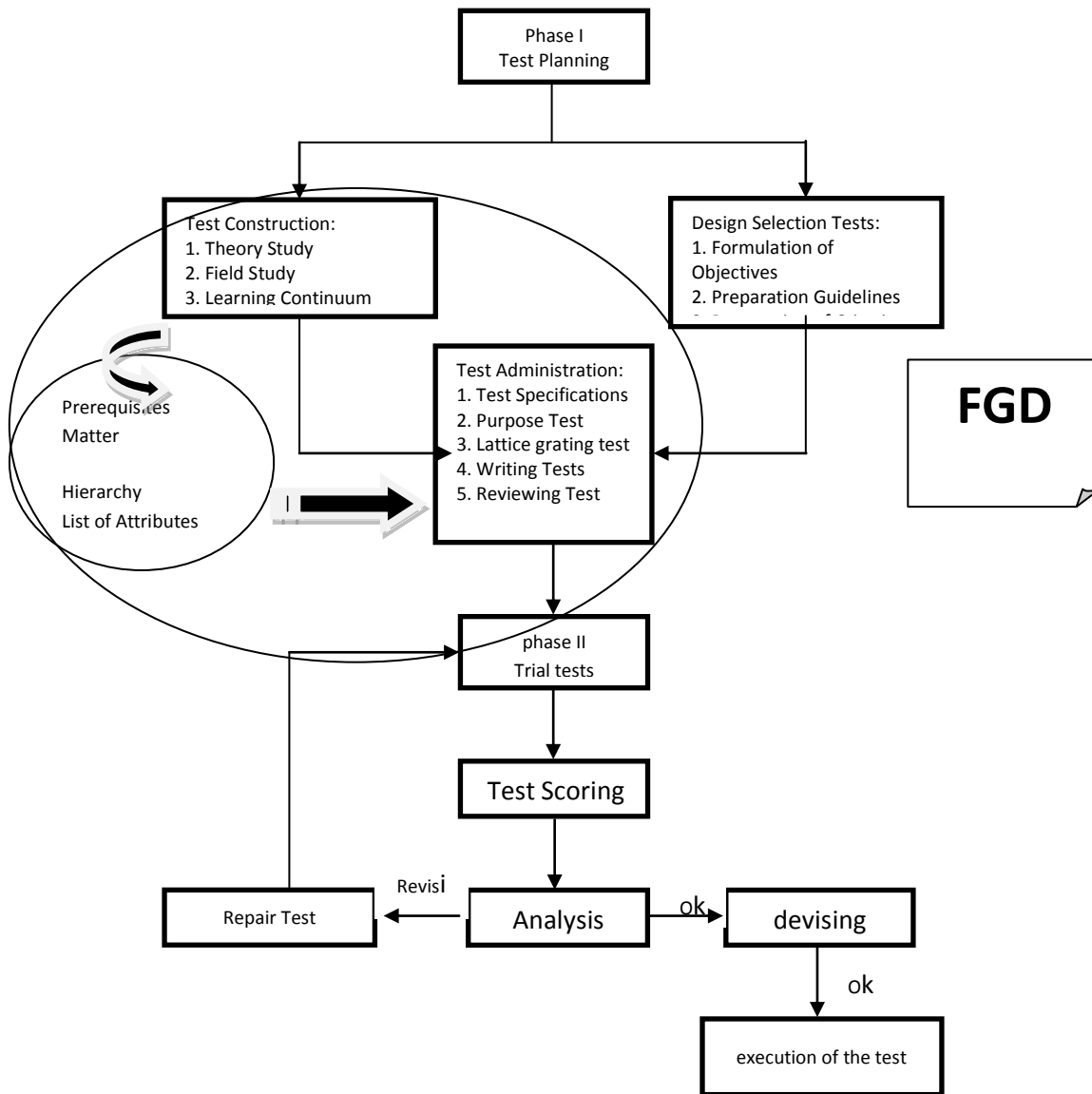


Figure 1. Research Desain

The research trials conducted in seven schools that have with high category, medium and low school category. Test subjects of the seventh grade students were 484. Development of data analysis based on descriptive cognitive diagnostic test done through Focus Group Discussion (FGD). To determine the quality of the instruments is done through validation empirically, valid item in the IRT models is used to assess the success of the calibration in an

effort to determine the fit of data to the model. If the item is declared fit to the model valid means of measurement (Wright & Masters, 1982: 114). Selection of test items to get fit tests was performed with polytomus Rasch Model, the Partial Credit Model (PCM) with the help of the Quest program. The Limit of one item is declared fit to the model if it has infit MNSQ between 0.77 up to 1:30 (Adam & Khoo, 1996). Fit statistics on program requirements quest, that is, if the value of the average infit MNSQ approaching 1.0 with a standard deviation of 0.00. If the data is shown to fit the models, means evident that all items measure the same construct (unidimensional).

Results and Discussion

The development of diagnostic tests instruments implemented in accordance with the lines of inquiry. At the planning stage the test starts with constructing substantive theories. At this stage also the formulation of the concept of mastery learning continuum materials science in SMP. Learning continuum that is made in order to describe the development and enhancement of the ability of learners to master the concepts of physics science started from the basic concepts to the rest of concept. Continuum formulation of learning through workshop that is followed by evaluation experts, education experts and 6 school teachers who are members of the IPA MGMPs Sleman having regard to the concept map. The next step is the selection of the design that is used to make items that can be responded construction either by candidates based on knowledge, specific skills or other characteristics according to the theory. At this stage of the draft prepared cognitive diagnostic instruments implemented in three activities, namely: the preparation of the formulation of objectives, preparation of guidelines for assessment and preparation of the assessment criteria. Develop test specifications (Item Specification) contains descriptions tests that show characteristics that must be possessed cognitive diagnostic tests. Preparation of test specifications include determining the purpose of the test, making the test grating, determine the shape of the test and determine the length of the test. The instrument is structured diagnostic tests on the material of heat by 28 items, with a selection of multiple choices answers created hierarchy.

Results of diagnostic tests of cognitive development in the first phase were analyzed by FGD. At this stage 18 followed by FGD participants consisting of measurement experts, cognitive psychology, science, education Physics, peers and junior high school teachers who are members of MGMPs Sleman. In FGD generating activities continuum Learning Subjects science in SMP calorific material consisting of 57 points, Heat Hierarchy material, test

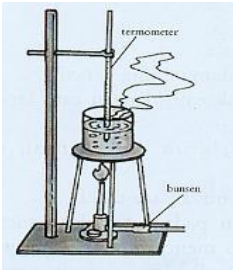
specifications, and test grating cognitive diagnostic test instruments in the form of multiple choices with reasons. Content validity of 4 experts is presented in the following table.

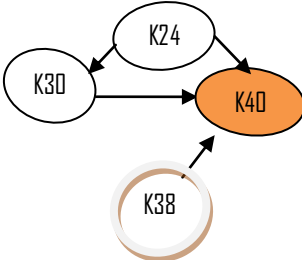
Table 1. Results of Focus Group Discussion

Contens	Not Good	Less	Good	Very Good
Learning Continuum	-	-	25%	75%
Prerequisites Matter Hierarchy	-	-	75%	25%
Grating test	-	-	50%	50%
Test specifications	-	-	-	100%
alternative Answers	-	-	25%	75%
Diagnostic information	-	-	50%	50%
Forms Questions	-	-	50%	50%
Answer sheet and answer key	-	-	-	100%

The overall instrument validation results the experts have good and excellent value. After the contents of the instrument validated by experts, then the instrument was tested in 484 junior high school students in seven schools in Sleman. Here's an example of cognitive diagnostic test specifications at number 16.

Table 2. Specification Tests in Item number 16

Type of School	: Junior High School	Allocation of Time	: 80 minutes (for all item)
Subject	: Sains	Form of Matter	: Multiple Choice by reason
Class	: VII	Author	: Team
Learning Continuum	: Explain the meaning of latent heat		
Type of Knowledge	: Factual		
Cognitive processes	: Comprehension		
Question	: Consider the following heat usage graph!		
			
	If the water temperature has reached 100 ° C and kept heated, then there is		

<p>Alternative Answers:</p> <p>A. The temperature of the water rises and the water mass is reduced B. The temperature of the water rises and the water condenses C. The temperature of the water remains and the water evaporates D. The temperature of the water rises and the water boils</p>
<p>Diagnostic information answer</p> <p>A. Answer incorrect Students can not master the material K 24 (a score of 2)</p> <p>B. Answer incorrect Students can not master the material K24 and K30 (a score of 1)</p> <p>C. The right answer (a score of 4) Students can master the material K40</p> <p>D. Answer incorrect (a score of 3) Students can not master the material K38</p>
<p>Image hierarchy Material prerequisite learning continuum</p>  <pre> graph TD K24((K24)) --> K30((K30)) K24((K24)) --> K40((K40)) K30((K30)) --> K40((K40)) K38((K38)) --> K40((K40)) style K40 fill:#f96 </pre> <p>Remarks:</p> <p>K24: Explaining the effect of heat in changing temperatures and states of matter K30: Mention the change occurring states of matter from liquid to gas K38: Distinguishing melting point, boiling point and vapor K40: Explain latent heat</p>

The data have been collected in empirically validate using PCM through the Quest program. Selection grains with the Quest program is used to get the value of item difficulty and test fit. In this study the number of items is 28 items, but in the process of doubling the instrument contained a typing error on the matter so that both questions and answer options are not appropriate. So the question is analyzed totaled 27. In this study, there is one person who answered the question correctly students all, if we follow the logistic curve, the students who received a score of correct all or any of all did not participate analyzed. so that the data analyzed in this study is data from trials sebyek by 483 students.

The overall results of the analysis with the Quest program are presented in the following table.

Table 3
 Estimation Results Item (i) Cognitive Diagnostic Test Subjects Physics Science and
 Estimated Student SMP (N) according to the Partial Credit Model Politomus

No	Information	item Estimates	Case Estimates
1	Mean & SD	0,01±0,54	0,97±0,40
2	Mean & SD (adjusted)	0,01±0,46	0,97±0,35
3	Reliability of estimate	0,71	0,74
4	Internal Consistency		0,71
5	Mean & SD infit MNSQ	1,01±0,09	1,02±0,30
6	Mean & SD outfit MNSQ	1,03±0,17	1,03±0,71
7	Mean & SD infit t	0,15±1,71	0,02±1,05
8	Mean & SD outfit t	0,16±1,71	0,12±0,77
9	cases with zero scores	0	0
10	cases with perfect scores	0	0

Note : N = 483 L = 27 Probability Level= .50

The Limit of one item is declared fit to the model if it has infit MNSQ between 0.77 up to 1,30 (Adam & Khoo, 1996). Requirements Fit statistics is on the Quest program, that is, if the value of the average infit MNSQ approaching 1.0 with a standard deviation of 0.00. Cognitive diagnostic tests instruments in science subjects have the number of items that were analyzed by 27 grains. From the results of the analysis showed that all items fit. In this study, all the items are in a vertical line or have infit MNSQ between 0.77 until 1.30. The research instrument is compliant statistics Fit on a quest program with an average value infit MNSQ 1.01 with a standard deviation of 0.09. Reliability values sampled in this study is 0.71. According to Wright & Masters (1982) the value of an item based on the estimated reliability is called reliability sample. Thus, the higher the value, the more items that fit with the model. While the reliability of the test in this study was obtained of 0.74 is by looking at the value of reliability based on the estimated case or testy. Reliability values based on the estimated case / test is called reliability test (Wright & Masters, 1982). The reliability of a test of 0.74 provides information that measurements using cognitive diagnostic test instrument on science lessons provide consistent results. The higher the reliability of the test, the more samples for tests that provide information that is expected, and vice versa. The lower of the value the less reliability of the test sample is to test the expected information. In this research, one student can answer all of the questions correctly. The level of difficulty of items in this study are presented in the following graph.

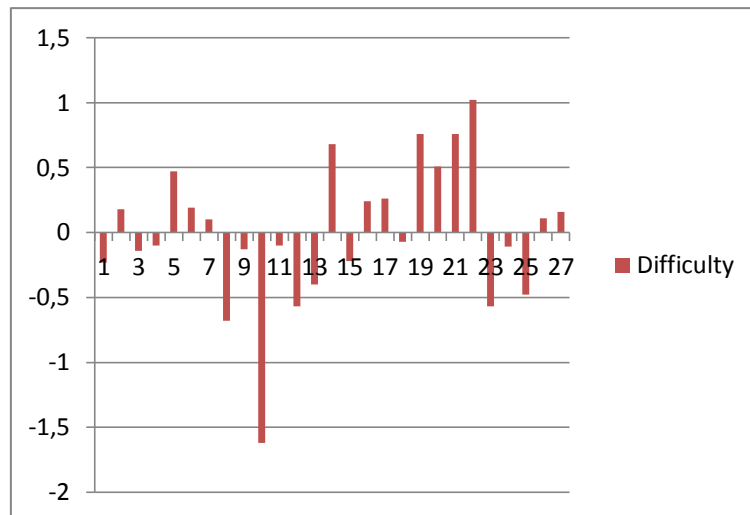


Figure 2.
Item Difficulty

Items good report if the index of difficulty (b) is between $-2.0 < b < 2.0$ (Hambelton & Swaminathan, 1985: 36). Based on the analysis, item difficulty of diagnostic tests of cognitive science subjects is in the interval -1.62 to 1.02. This means that as many as 27 items of diagnostic tests of cognitive science this is all good. In the figure it can be seen that the most difficult problem is the question about the number 22 and the easiest is the matter of the number 10.

Conclusion

The study successfully prepared learning continuum, grating test, test specifications and develop cognitive diagnostic tests instruments in the form of multiple choices with reasons. The options granted are based on a hierarchy of prerequisite material. Tests were successfully developed consisting of 28 items, and which can be analyzed is 27 items. By using gained 27 Quest program items that fit by looking that the items are in a vertical line or have infit MNSQ between 0.77 until 1.30. The reliability value of 0.74 tests provides information that measurements using cognitive diagnostic tests instruments in science lessons provide consistent results. The research instrument is compliant Fit statistics on the Quest program, which is the average value of infit MNSQ 1.01 with a standard deviation of 0.09.

References

Adam J. Raymond & Khoo Siek-Toon. (1996). *Quest. The Interactive Test Analysis system*: The Australian council For Educational Research.

- Ebel, Robert, L. (1979). *Essentials of Educational Measurement*. New Jersey: Prentice-Hall, Inc.
- Efendi, Ridwan. (2010). *Kemampuan Fisika Siswa Indonesia Dalam TIMSS (Trend of International on Mathematic and Science Study)*. Prosiding Seminar Nasional Fisika 2010. ISBN : 978-979-98010-6-7
- Embretson . (2002). *Generating abstract reasoning item with cognitive theory*. In Irvine, S.H., & Kyllonen, P.C. *Item generation for test development* (pp 219-250). Mahwah, N.J.: LEA Publisher
- Gorin, J.S. (2007). *Test construction and diagnostik testing*. In Leighton, JP. & Gierl, M.J. (eds). *Cognitive diagnostik assessment for education theory and applications*. (pp 174). Cambridge. Cambridge University Press.
- Hambelton & Swaminathan. (1985). *Item Response Theory Principles an Application*. Boston. Kluwer Nijhoff Publishing.
- Hungerford, D.K. & Volk. & Ramsey. (1986). *Science-Technology-society*. New York : McGraw-Hill.
- Keeves, J.P. & Masters, G.N. (1999). *Introduction* In G.N Masters & J.P Keeves (Eds), *Advances in Measuremen in Educational research and assessment* Amstredam: Pergamon An imprint of Elsevier Science.
- Mehrens, W.A & Lehman, I.J. (1973). *Measurement and evaluation in education and Psychology*. New York. Holt, Rinehart and Winson, Inc.
- Nichols. (1994). *A framework for developing cognitively diagnostik assessment*. Review of Educational . 64 (4), 575-603)
- Nichols, PD. Chipman, S.F and Brennan, R.L (1994). *Cognitively diagnostik Assessment*. New Jersey: Lawrence Erlbaum Associates
- Titus, H.H. (1959), *Living Issues In Philoshophy*, New York: American Book Company.
- Weeden, P. Winter, J. & Broadfoot, P. (2002). *Assessment what's in for school?*. Routledge falmer: London
- Wright, B.D & Masters, GN. (1982). *Rating Scale analysis*. Chicago: Mesa Press.