

TEST ITEM ANALYSIS PROGRAM DEVELOPMENT WITH RASCH MODEL ONE PARAMETER FOR TESTING THE ITEM DIFFICULTY LEVEL OF MULTIPLE-CHOICE TEST USING BLOODSHED DEV C ++ APPLICATIONS

Dadan Rosana², Otok Ewi Amsirta¹

¹Lecturer in Physics Education Studies Program, Yogyakarta State University,
email: danrosana.uny@gmail.com

²Students of Physics Education Study Program, Yogyakarta State University

Abstract

This article was made in Item Response Theory lecture to improve the ability of students to learn the application of the theory of grain analyst using the model of Rasch model. This article is aimed to get a test item analysis program with Rasch model one parameter for testing the item difficulty level of multiple-choice test and to determine the distribution of the item difficulty level of the test which is analyzed using the result program of the development. The development of test item analysis program were performed using 4D models which is consist of define, design, develop and disseminate. The results of program development named RASCHAM. RASCHAM using Item Response Theory (IRT) which adapt the Rasch model one parameter. Based on the result of validation from four validators, the test item analysis program judged worthy used for test item analysis. Based on testing performed by comparing the analysis results of the QUEST can be concluded that the accuracy of RASCHAM program reached 92.80%. The analysis results of the item difficulty level with RASCHAM on the odd semester of final examination test items of the physics subjects from XI class of 2th Wonosari State Senior High School academic year 2013/2014 in Gunungkidul for code A is about 7.5% with a very easy category, 12.5% easy, 25% moderate, 47.5% difficult, 2.5% very difficult and 5% not good. While for code B is about 12.5% with a very easy category, 32.5% easy, 45% moderate, 45% difficult, 2.5% very difficult and 2.5% not good.

Keywords: *Rasch Model, Item Response Theory (IRT), RASCHAM, QUEST*

Introduction

The quality of education can not be separated from the quality of the performance of the teachers. That's why educators and professionals are required to have extensive knowledge and insight to be transferred to the learners. Professionalism of educators can be seen from the professionalism in carrying out tasks anyway. One of the main tasks of educators is to conduct an evaluation of the education process.

Discusses the evaluation of education recognized the existence of 8 quality assurance standards of education including content standards, process standards, competency standards, teachers and standards, standards of infrastructure, management standards, financial standards, and assessment standards. Standard assessment is an evaluation to measure student learning outcomes as prestasi learning. This means that the acquisition should be in accordance with the subject matter competency. Evaluation of learning outcomes is a series

of systematic and ongoing process to determine the quality of the learning that is based on certain predefined criteria. According to Budi Susilowati Emi (2012: 1), the evaluation of teaching and learning activities are also regulated in Law No. 20/2003 on National Education System in paragraph 1 which states that the evaluation is done in order to control the quality of education nationwide as a form of accountability of education providers to the parties concerned.

So far, there has not been an evaluation based on the analysis of good items. Analysis of items is a term used to define calculations and measurements of the subject's response to an item (Crocker & Algina, 1986). In general, item analysis aims to determine whether an item is an item that is good or bad as a measuring instrument. Analysis items are basically divided into two categories: analysis of qualitative and quantitative questions.

Qualitative analysis of a study intended to analyze the problem in terms of technical, content, and editorial. Quantitative analyzes point is a review of items based on the empirical data of the item in question. The empirical data obtained from the questions that have been tested. There are two approaches in the quantitative analysis, the classical and modern approaches.

Classical item analysis is the review process items through information from the answers of students in order to improve the quality of items is concerned with the use of classical test theory. (Millman and Greene, 1993: 358). Aspects to be considered in the analysis of classical item is any item assessed in terms of: level of difficulty grains, grain distinguishing features, and deployment of answer choices (for a matter of objective shape) or the frequency response at each answer choice.

Analysis of modern items that the review items using Item Response Theory (IRT) or item response theory. This theory is a theory that uses mathematical functions to connect the right opportunity to answer a question with a student's ability.

The use of classical item analysis there are still weaknesses. Weaknesses include the item parameters and the parameters are interdependent so that participants can not be generalized to other groups of participants. While the model of Item Response Theory (IRT) was able to cover the weaknesses of classical models. In the model of Item Response Theory (IRT) no dependence on grain parameters and parameters of the participants.

During this time the program has been developed using either item analysis of classical models and the model of Item Response Theory (IRT). One program of quantitative analysis of the popular items used are Iteman. Iteman developed by Micro Computer Adaptive Test (MICROCAT) Assessment Systems Corporation, the University of Pittsburgh. The program is very simple so that by studying a moment would've been able to master it. However Iteman

programs still use the classical theory, so the consumer still has weaknesses. The program can be used to analyze items to the model Item Response Theory (IRT), among others: Rascal, Pascal, Bigsteps and Quest. Each program has a different procedure to run. However, existing procedures it is still too difficult to be understood, so that the programs that should be used to facilitate analysis of items even harder because the procedure is too complicated and long. Programs item analyzes both quantitative analysis of the classical model and the model of Item Response Theory (IRT), which has been developed and disseminated largely dominated by foreign-made. While the program being developed in the country is still lacking. Untukitu, the author had the idea to develop a program analysis using the model item Item Response Theory (IRT). The model of Item Response Theory (IRT) used is specific to one parameter Rasch model. Program development results are validated by comparing the results of the development program with existing programs (QUEST).

Methods

This study aimed to obtain item analysis program with one parameter Rasch models for testing achievement test devices using Bloodshed Dev C ++ is feasible and accurate. Development model used in this study is a model 4D (Four-D Model).

Model 4D (Four-D Model) consists of defining phases (define), the design phase (design), stage of development (develop), and the dissemination phase (disseminate). The stages of the 4-D models are described as follows:

1. Defining Phase (define)

Defining in this case is to establish and define the needs in the development of item analysis program. Things to consider is the formatting and program development techniques. Through this defining stage formatting and program development techniques are analyzed based on their needs and in accordance with the criteria of modern item analysis.

2. Stage Design (design)

The objective of the design phase is to design a format item analysis program with one parameter Rasch models for testing the test results belajar menggunakan Bloodshed Dev C ++. In this phase, the program format item analysis using item analysis techniques one parameter Rasch models. Development is done using Dev C ++ software Bloodshed. The programming language used is C ++.

3. Development Phase (develop)

At this stage the program was developed with Rasch item analysis of the parameters of the model for testing the achievement test using Bloodshed Dev C ++ and ready for

use. Furthermore, in this stage of the evaluation and analysis of the program revision on items that have been made. Evaluation is done by consulting programs that have been made to two expert lecturers and 2 users with the intent to obtain advice. Then, be revised in accordance with the advice given by the validator.

4. Dissemination (disseminate)

At this stage, the dissemination of product development efforts. Dissemination of product development is done in SMA N 2 Wonosari by way of disseminating product development is limited only to teachers of physics.

Results and Discussion

The results of the research in the early stages of defining activities include establishing and defining the needs of the development program items with the Rasch analysis of the parameters of the model for testing the quality of the test results to learn multiple choices. Things to consider is the formatting and program development techniques. Through this defining stage formatting and program development techniques are analyzed based on their needs and in accordance with the criteria of modern item analysis.

To meet the criteria for the analysis of the modern items used Rasch item analysis techniques parameter. Pengertian model of one of the parameters here are the result of analysis obtained by the analysis of item. So the definition of the parameters is not a criterion of input data to be processed. The input data to be processed is a student answers a multiple choice. This input data is free in the sense that is not tied to a particular matter, the basic competence (KD) specific, or specific goals.

The objective of the design phase is to design a format item analysis program with one parameter Rasch models for testing the achievement test. At the design stage researchers collect references that support the development of program analysis items. References obtained came from a book, article or journal from the internet.

After references collected, researchers began to draft a plan and analyze it. The results of the draft plan include item analysis techniques; use software to create a program; the use of a programming language in making the program; making of the program flow scheme in outline; and prototype RASCHAM program.

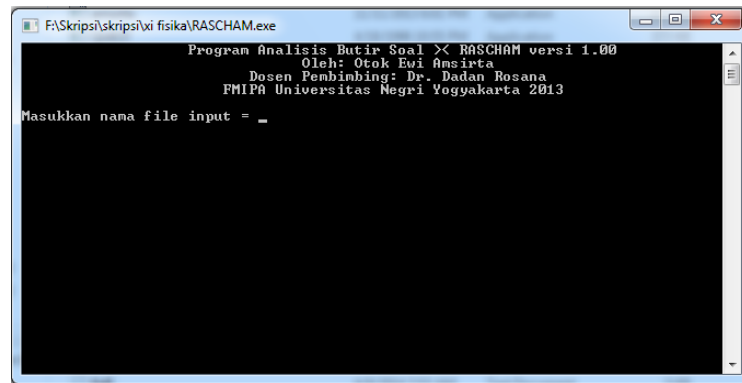


Figure 1. Display Program

At this stage of development has been done consists of the manufacture of items RASCHAM program analysis, validation RASCHAM, revised results RASCHAM validation, testing RASCHAM program. Validation program conducted by validator 2 and 2 users. In Tabel 1 and Table 2 respectively presented the results of the average ratings by the media RASCHAM validator and test results.

Table1. Recapitulation Item Difficulty Level for Code A Problem With Using RASCHAM Program

| Category | Item | Numbers | Percentage |
|--|---|-----------|------------|
| very easy ($-3 \leq bi \leq -2$) | 10, 17, 18 | 3 | 7.5 |
| easy ($-2 \leq bi \leq -1$) | 4, 16, 32, 33, 36 | 5 | 12.5 |
| moderate ($-1 \leq bi \leq 0$) | 2, 3, 11, 14, 19, 20, 21, 23, 27, 39 | 10 | 25 |
| difficult ($0 \leq bi \leq 2$) | 1, 5, 6, 7, 12, 13, 15, 22, 25, 26, 28, 29, 30, 31, 34, 35, 37, 38, 40 | 19 | 47.5 |
| very difficult ($2 \leq bi \leq 3$) | 24 | 1 | 2.5 |
| not good ($-3 > bi$ or $3 < bi$) | 8, 9 | 2 | 5 |
| Total | | 40 | 100 |

This study aimed to obtain an item analysis program with one parameter Rasch models for testing the quality of the multiple-choice achievement test (multiple choices) .Pembuatan soal ini item analysis program has dibuat sesuai with the planned design. The developed program is named RASCHAM program. The look of the program are as follows RASCHAM.

Table 1. Recapitulation Item Difficulty Level for Code B Problem With Using RASCHAM Program

| Category | Item | Numbers | Percentage |
|--|---|-----------|------------|
| very easy ($-3 \leq bi \leq -2$) | 5, 12 | 2 | 12.5 |
| easy ($-2 \leq bi \leq -1$) | 3, 9, 11, 31, 37 | 5 | 32.5 |
| moderate ($-1 \leq bi \leq 0$) | 2, 7, 8, 13, 15, 16, 20, 22, 25, 27, 28, 34, 38 | 13 | 45 |
| difficult ($0 \leq bi \leq 2$) | 1, 6, 10, 14, 17, 18, 19, 21, 23, 24, 26, 29, 30, 32, 33, 35, 36, 40 | 18 | 45 |
| very difficult ($2 \leq bi \leq 3$) | 39 | 1 | 2.5 |
| not good ($-3 > bi$ or $3 < bi$) | 4 | 1 | 2.5 |
| Total | | 40 | 100 |

The analysis showed that item difficulty index replicates the final semester of high school physics subjects N 2 Wonosari class XI Science in Gunung Kidul 2013/2014 school year about a code stretches from -2.13922 to 2.12202 while the code is about B extends from -2.82609 to 2.49759 . For Problem A code indicating that the level of difficult questions that have as much as 2.13922 which is about the number 10, 17, and 18 is about the easiest, while the matter of having a difficult level of 2.12202 is about the number 24 is the most difficult problem. For about the code bahwas oal B shows that have a level of difficulty of -2.82609 is about number 5 and 12 are about the easiest. As for who has the difficult level 2.249759, ie number 39 is the most difficult problem.

Conclusion

Based on the analysis of program validation of all aspects of the item in the excellent category by faculty experts and well by all three programs other validators, so that the item analysis program used for the analysis of feasible items. Based on tests carried out by comparing the results of the QUEST program analysis can be concluded that the accuracy reaches 92.80% RASCHAM program.

The results point to the difficulty level of analysis RASCHAM program on items Deuteronomy End Semester (UAS) odd subjects in class XI physics SMA N 2 Wonosari 2013/2014 school year in the district of Gunung for Problem A code is about 7.5% with a very easy category, about 12.5% easy, 25% moderate problem, 47.5% about the difficult, very difficult about 2.5% and 5% did not matter either. As for the matter of code B is about

12.5% with a very easy category, 32.5% easy matter, about 45% moderate, 45% about the difficult, very difficult about 2.5% and 2.5% does not matter either.

References

- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Forth Worth: Holt, Rinehart, and Winston, Inc.
- Dali S Naga. (1992). *Pengantar Teori Sekor pada Pengukuran Pendidikan*. Jakarta: Gunadarma
- Depdiknas. (2008). *Panduan Analisis Butir Soal*. Jakarta: Depdiknas
- _____. (2009). *Penilaian*. Jakarta: Depdiknas
- Djemari Mardapi. (2004). *Penyusunan Tes Hasil Belajar*. Yogyakarta: UNY
- Emi Budi Susilowati. (2012). *Pengembangan Sistem Evaluasi Kegiatan Belajar Mengajar Berbasis Web*. Yogyakarta: UGM
- Gronlund, N.E., Linn, R.L., & Miller, M.D. (2009). *Measurement and Assessment in Teaching*. Upper Saddle River: Merrill
- Hambelton, R.K., & Swaminathan, H. (1985). *Item Response Theory Principles and Applications*. Boston: Kluwer
- Hambelton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. California: Sage Publications, Inc.
- Insap P. Santosa. (1994). *Grafika Komputer dan Antarmuka Grafis*. Yogyakarta: Penerbit Andi Offset
- James S. Cangelosi. (1995). *Merancang Tes untuk Menilai Prestasi Siswa*. Bandung: Penerbit ITB
- Millman, Jason and Greene, Jennifer. (1993). The Spesification and Development of Tests of Achiievement and Ability in Robert L. Lin (Editor). *Educational Measurement*, Third Edition. Phoenix: American Council on Education, Series on Higher Education Oryx Press
- Mundilarto. (2010). *Penilaian Hasil Belajar Fisika*. Yogyakarta: P2IS FMIPA UNY
- Nana Sudjana. (2011). *Penilaian Hasil Proses Belajar Mengajar*. Bandung: Remaja Rosdakarya
- Nana Syaodih Sukmadinata. (2009). *Metode Penelitian Pendidikan*. Bandung: Rosdakarya
- Punaji Setyosari. (2012). *Metode Penelitian Pendidikan dan Pengembangan*. Jakarta: Kencana Prenada Media Group
- Purwanto. (2011). *Evaluasi Hasil Belajar*. Yogyakarta: Pustaka Pelajar

Reynolds, C.R., Livingston, R.B., & Willson V. (2009). *Measurement and Assessment in Education*. New Jersey: Pearson Education, Inc.

Suharsimi Arikunto. (2009). *Dasar – Dasar Evaluasi Pendidikan (Edisi Revisi)*. Jakarta: Bumi Aksara

Trianto. (2007). *Model-Model Pembelajaran Inovatif Berorientasi Konstruktivistik*. Jakarta: Prestasi Pustaka

Zaenal Arifin. (2009). *Evaluasi Pembelajaran*. Bandung: Remaja Rosdakarya