

Application of the Rasch Model in Analysis of Exam Questions at the Faculty of Psychology of Universitas Islam Bandung

Dwi Agustin Nuriani Sirodj*

Program Study of Statistics
Universitas Islam Bandung
Bandung, Indonesia

*dwi.agustinnuriani@unisba.ac.id

Rizka Hadian Permana, Suhana Suhana

Faculty of Psychology
Universitas Islam Bandung
Bandung, Indonesia

Abstract—In knowing student achievement in a particular course, students often use examinations conducted at least twice a semester. It is expected to be able to measure accurately the extent of student competence in the course concerned. But in reality, not all students get good quality grades. However, the first thing to be ascertained when student achievement is low is a measuring tool used to measure learning achievement itself whether it is already of good quality or not. When the results obtained come from a low quality achievement test kit, then the wrong conclusions will be obtained. The method used to analyse the quality of the performance measuring instrument is the Rasch model. Based on the research that has been done on the exam questions at the Faculty of Psychology Universitas Islam Bandung which is for Neuropsychology courses it is found that the quality of the exam questions is not good. This can be seen from the number of problem items that need to be revised in total or replaced with an average of 50%. The number of questions that need to be replaced illustrates that these questions cannot distinguish students who have high ability from students who have low ability. The low ability of questions to distinguish the ability of respondents in dealing with problems can be caused by the unequal level of difficulty of the item problem.

Keywords: *test questions, achievement tests, Rasch model*

I. INTRODUCTION

The quality of education of a tertiary institution is determined by the quality of its graduates. Achievement of graduate quality can be seen from the cumulative achievement index (GPA) obtained by students while studying at the tertiary institution. The GPA is a picture of the learning achievements taken by students during the teaching and learning process in college. According to Winkel, learning achievement is a change in cognitive, sensory-motoric and dynamic-affective field, which can make a person change his attitude and behaviour. Changes in the cognitive field occurs when someone changes from not knowing to knowing something [1].

A concept that is very well known in the field of Education and is often referenced to identify thinking skills ranging from low levels to high levels is the concept of Bloom's Taxonomy. In Effendi explained that Benjamin. S. Bloom makes a classification based on the order of thinking skills in a process

that is getting higher to the higher levels [2]. At first bloom taxonomy consists of two parts, namely the cognitive domain and the affective domain. In 1966, Simpson added the psychomotor domain to complement what bloom had made. Thus, the three domains are cognitive, affective and psychomotor. However, along with the times, in Andeson and Krathwohl, revisions were made in the cognitive domain into six levels [3], namely: remembering, understanding, applying, analysing, evaluating, and creating.

For undergraduate students who have high levels of thinking skills, if they follow the taxonomic stage Bloom above, they have entered the evaluating stage. To measure the extent of knowledge or understanding that has been acquired after participating in learning activities, in general, tests will be conducted which are often called midterms or end of semester exams.

At the Faculty of Psychology, Universitas Islam Bandung, to find out student achievement in a particular course, it often uses exams that are conducted at least twice in one semester. It is expected to be able to measure accurately the extent of student competence in the course concerned. All Psychology faculty students are certainly expected to get satisfactory grades in all courses by obtaining a minimum quality letter B or average GPA of 3.0. However, in reality, not all students get good quality grades. There are some specific subjects which generally get satisfactory grades below. Based on this, it can be said that student achievement in several subjects is low. However, in the first time, it must be ensured that low student achievement is a tool to measure learning achievement itself whether it is in a good quality or not. The tool that is generally used by lecturers in the Faculty of Psychology is achievement tests or examinations. The tests / exams given are in the form of problems given to students which can be given from various forms of essays, multiple choice, short entries, matchmaking, and so forth.

The quality of an item in a test can be analysed both qualitatively and quantitatively. Qualitatively, it can be seen from the content of the problem given whether it is in accordance with the Course Implementation Plan or the syllabus of the course. In the Faculty of Psychology, the

content of lecturers in the course has been arranged according to the course implementation plan or syllabus of the relevant subject. So that, the problems given are in line with what students have studied in previous classes. As for the quantitative analysis, it can be done with a statistical analysis approach. Research on evaluating the quality of exam questions at the Faculty of Psychology has been carried out in Borualogo for Pedology courses, it's just that the method used is still using the classical test theory (CTT) method [4]. Another method that can be used is the Rasch Model. The Rasch Model is one of method to assess the feasibility of the problem by comparing the ability of respondents with the opportunity of correct answers in the group predicted by the model. One of the features of the Rasch Model does not depend on the sample used [5]. Rasch model is one of the models in the item response theory (IRT) method, namely the One Parameter Logistics Model (1PL). IRT is a theory of how person variables and item variables determine response data when someone answers the item [6]. Comparison of IRT and CTT methods is also thoroughly examined to determine which method is best in assessing the quality of an item. The advantage of the IRT method compared to the CTT method is that the item statistics do not depend on the group, the test scores obtained can describe individual abilities, it does not require parallel tests to calculate the reliability coefficient, and it can provide the right size for each ability score [6]. So it can be said that the IRT method is better in assessing the quality of the item than the CTT method.

Rasch model is widely applied to evaluate psychological measurement tools such as personality disorders [7], patient-rated wrist evaluation [8], Hospital Anxiety and Depression Scale (HADS) [9], Depression Anxiety and Stress Scale [10], and bullying [11] and so on. Besides the Rasch model is also used to evaluate examination questions in mathematics course [12].

By using this model, we can find out that students who have a higher ability will have a greater probability of answering questions correctly. To test the test / test item quantitatively, it is rarely done by lecturers in the Faculty of Psychology, Universitas Islam Bandung. On this basis the researcher wants to find out the quality of the items on tests / examinations given to psychology students, especially subjects whose students generally get below satisfactory grades using the Rasch Model analysis.

II. METHODOLOGY

A. Participants

The population of this research was the whole active students of Faculty of Psychology of Universitas Islam Bandung year of 2018-2019.

B. Data Collections and Analysis

In this research, it is certainly not all exam questions in all courses that take place at the Faculty of Psychology will be analysed. Since the 2017-2018 academic year, the Faculty of Psychology has implemented a new curriculum called the 2017

Higher Education Curriculum (Kurikulum Perguruan Tinggi (KPT)).

The process of determining what subjects will be sampled is carried out by the judgmental sampling method. The subjects that will be analysed for the exam questions are courses that are based on previous data that were obtained from students who get final grades C, D and E exceeding 20%.

III. RESULTS AND DISCUSSION

A. Determination of Sample Test Questions

The first stage of the analysis is to use secondary data about the distribution of student grades in each subject. Following is a recapitulation of the distribution of value data obtained by students in each course held in the even semester of the 2018-2019 academic year for the KPT curriculum.

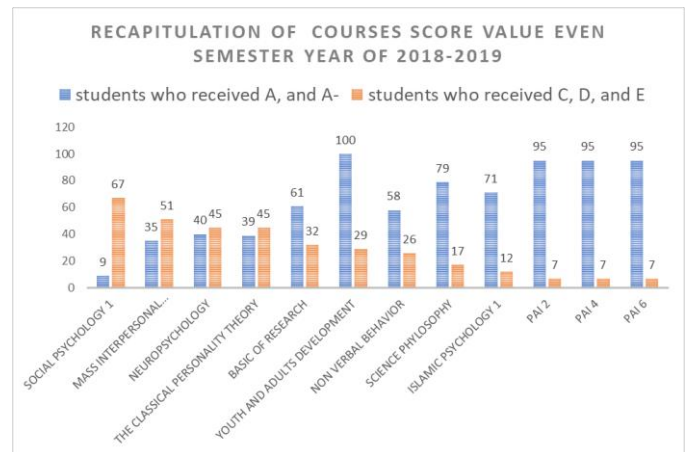


Fig. 1. Recapitulation of courses score value even semester year of 2018-2019.

Based on the recapitulation in Figure 1, the KPT courses in the even semester that will be sampled and analysed are exam questions for Social Psychology 1/Mass Interpersonal Communication/Neuropsychology courses because these three courses has the most students who received C, D, and E.

After determining which subjects will be analysed for exam questions, the next process is identifying the type of exam questions, and checking the availability of student exam files so that answers can be input, but from several predetermined courses obtained the type of exam questions is in the form of essays. So that, it cannot be analysed further, besides the unavailability of exam files from lecturers also become an inhibiting factor in collecting data. Thus, it was determined that another course to be analysed for the exam questions was Neuropsychology.

B. Analysis of Neuro-Psychology Subjects

1) Part I

TABLE I. RECAPITULATION OF INDEX FACILITY NEUROPSYCHOLOGY QUESTIONS PART I

Index Facility Criteria	Number of Questions
Enough	14
Too easy	10
Too difficult	6
Total	30

Based on Table I above, 47% of the Neuropsychology exam section 1 is on a level of difficulty that is classified as sufficient to be answered by respondents, 33% of problems have difficulty levels on “too easy” and 20% of problems have levels of problem on “too difficult”.

TABLE II. RECAPITULATION OF INDEX DISCRIMINATION NEUROPSYCHOLOGY QUESTIONS PART I

Index Discrimination Criteria	Number of questions
Good	2
Enough	6
Eliminated/Revise Total	13
Less Revise	9
Total	30

Based on table II above, 43% of neuropsychological exam questions for Part 1 must be totally revised or replaced, as many as 30% of the problems should be slightly revised and as many as 6% of the problems that are already good. In general, both neuropsychological problems in section 1 cannot distinguish students who have high abilities from students who have low abilities.

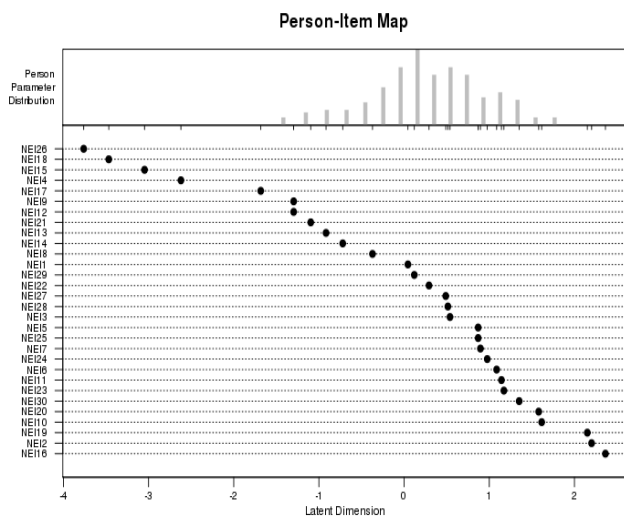


Fig. 2. Person-item map neuropsychology questions part I.

Based on Figure II above, it can be seen that there are 4 neuropsychological problems, namely questions number 26, 18, 15, and 4 which are too easy and far below the normal criteria that move from the category -2 to 2. When it is viewed from the uneven distribution, it relatively accumulates in the

middle so as to make there are quite a number of issues that need to be replaced and revised.

2) Part II

TABLE III. RECAPITULATION OF INDEX FACILITY NEUROPSYCHOLOGY QUESTIONS PART II

Index Facility Criteria	Number of questions
Enough	12
Too difficult	8
Total	20

Based on table III above, 60% of the Neuropsychology test section 2 has a level of difficulty that is classified as sufficient to be answered by respondents, 40% of problems have ‘too difficult’ and 0% of problems have ‘too difficult’.

TABLE IV. RECAPITULATION OF INDEX DISCRIMINATION NEUROPSYCHOLOGY QUESTIONS PART II

Index Discrimination Criteria	Number of questions
Good	6
Enough	1
Eliminated/Revise Total	11
Less Revise	2
Total	20

Based on table IV above, 55% of neuropsychological examination questions part 2 must be totally revised or replaced, as many as 10% of the problems should be slightly revised and as much as 10% of problems that are already good. In general, both neuropsychological problems in part II cannot distinguish students who have high ability from students who have low ability. A reward for power in bullying was expressed on the amount of bully. The more students get a part of bully, the more powerful they are.

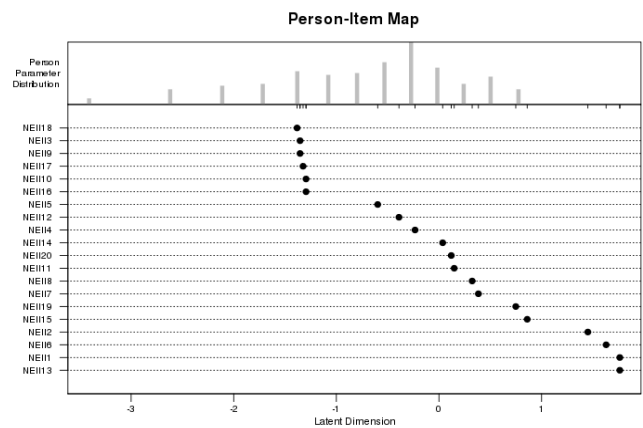


Fig. 3. Person- item map neuropsychology questions part II.

Based on figure III above, if it is seen from the unequal distribution relative to pile up, it indicates that only a small number of respondents have the opportunity to answer correctly on the issues presented, making the number of issues to be replaced and revised quite a lot.

IV. CONCLUSION

Based on research that has been done on Neuropsychology subject matter tests, it is found that the quality of the subject exam questions is not good. This can be seen from the number of problem items that need to be revised in total or replaced with an average of 50%. The number of questions that need to be replaced illustrates that these questions cannot distinguish students who have high ability from students who have low ability. The low ability of questions to distinguish the ability of respondents in dealing with problems can be caused by the unequal level of difficulty of the item problem. Like the results of the analysis conducted on the three subjects, it was found that the unequal distribution of the degree of difficulty of the item. On average 50% of the problems are in the very easy category, so the chances of respondents both high and low ability to answer the problem correctly are equally high.

REFERENCES

- [1] W. Winkel, Psikologi Pengajaran, Jakarta: PT Media Abadi, 2009.
- [2] R. Effendi, "Konsep Revisi Taksonomi Bloom Dan Implementasinya Pada Pelajaran Matematika SMP," *Jurnal Ilmiah Pendidikan Matematika*, vol. 2, no. 1, pp. 72-28, 2017.
- [3] L.W. Andeson and D.R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assesing: A Revision of Bloom's Taxonomy of Educatioanl Objectives.*, New York: Addison Wesley Longman, Inc. , 2001.
- [4] I.S. Borualogo, S. Kusdiyati, Susandari and D.A.N. Sirodj, "Analisis Item Soal Uts Pedologi Semester Ganjil 2015-2016," *SCHEMA - Journal of Psychological Research*, pp. 46-57 , 2017.
- [5] K.D. Kubinger, D. Rasch and T. Yanagida, "A new approach for testing the Rasch model," *Educational Research and Evaluation*, vol. 17, no. 5, p. 321-333, 2011.
- [6] D. Huriaty, "Metode Kalibrasi Dan Desain Tes Berdasarkan Teori Respon Butir (IRT)," *Math Didactic : Jurnal Pendidikan Matematika Vol 1, No 3*, 2015.
- [7] L.d.F. Carvalho, R. Primi and G. J. Meyer, "Application of the Rasch model in measuring personality disorders," *Trends Psychiatry Psychother*, vol. 34, no. 2, pp. 101-109, 2012.
- [8] S. Esakki, J.C. MacDermid, J.I. Vincent, T.L. Packham, D. Walton and R. Grewal, "Rasch analysis of the patient-rated wrist evaluation questionnaire," *Archives of Physiotherapy* , vol. 8, no. 5, 2018.
- [9] J.F. Pallant and A. Tennant, "An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS)," *British Journal of Clinical Psychology* , vol. 46, pp. 1-18, 2007.
- [10] T.L. Shea, A. Tennant and J.F. Pallant, "Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS)," *BMC Psychiatry*, vol. 9, no. 21, pp. 1-10, 2009.
- [11] L. Kyriakides, C. Kaloyirou and G. Lindsay, "An analysis of the Revised Olweus Bully/Victim Questionnaire using the Rasch measurement model," *The British Psychological Society* , pp. 781-801, 2006.
- [12] Z.M. Nopiah, M.H. Jamalluddin, N.A.I.H. Othman, I. Asshaari and M.H. Osman, "Reliability Analisis of Examination Questions in a Mathematics Course Using Rasch Measurement Model," *Sains Malaysiana* , vol. 41, no. 9, pp. 1171-1176, 2012.