C9 Quality analysis of final test in the faculty engineering of Yogyakarta State University

by Emy Budiastuti

Submission date: 18-Jun-2020 12:09PM (UTC+0700)
Submission ID: 1345779149
File name: st_in_the_faculty_engineering_of_Yogyakarta_State_University.pdf (687.47K)
Word count: 2097
Character count: 11481

Quality analysis of final test in the faculty engineering of Yogyakarta State University

Quality analysis of final test in the faculty engineering of Yogyakarta State University

E Budiastuti[1], Sugiyono[2], and MA Jerusalem[3]

[1,2,3]Faculty of Engineering, Yogyakarta State University, Yogyakarta, Indonesia

E-mail: emy_budiastuti@uny.ac.id

Abstract. Final test is conducted to assess ability and skills within particular time. Before students take the test, the questions must qualify for good test. The previous research in Hospitality and Fashion Education Study Program reveals that the lecturers have not completely developed final test questions based on required principles of composing questions. This research aims to: 1) identify the quality of final tests questions of theory course, and 2) identify the quality of final tests questions of practice course including questions' validity and reliability. This is a survey research. The population of this research was all final test questions in the Faculty of Engineering. Meanwhile, the sample of this research was final test questions of theory and practice courses. 63 samples were collected through purposive sampling technique. The quality of final test questions of theory course was analysed by employing Anates while the questions of practice course was analysed by employing Kappa. The data were analysed by employing descriptive analysis. The results indicate that: 1) difficulty level of multiple choices and essay questions is fair; 2) discrimination power of the questions is fair; 3) effectiveness of tricky questions is poor; 4) item validity of multiple choice and essay is fair; 5) reliability of multiple choice questions and theory is poor; 6) validity of questions of practice course is good; and 7) reliability of the questions is good with reliability index is > 0.7.

1. Introduction

Each learning process taught by lecturers must have a test. There are two types of tests: mid-term test and final test. The tests are applied to investigate students' achievement level during the learning process. Final test is one of activities evaluating students' learning achievement, and it is particularly conducted in the Faculty of Engineering at Yogyakarta State University. Evaluation is scoring students' ability to receive, comprehend, and master courses taught based on predetermined curriculum, as well as assessing change in their behaviour and skill. The result of previous research indicates that most of final test questions, particularly in theory courses, are not based on principles of composing good questions. To investigate the quality of test questions developed by lecturers,item analysis is required. Final test aims to 1) investigate students' level of learning achievement including cognitive, affective, and psychomotor aspects within particular time; 2) investigate effectiveness of learning process; and 3) determine level of learning achievement in theory course into five categories: very good, good, fair, poor, and very poor, as well as in practice course into two categories: competent and not competent. To determine the categories, guideline and assessment rubric are required. Rubric is composed to prevent assessor's subjectivity and to achieve reliability level ofinter-raters(Bresciani, 2009: 2-3).

Performance assessment is a process of collecting information through systematic observation to determine policy on individual (Berk, 1986:ix). Smith (2007:2) posits that in performance test, inter-rater reliability is possibly employed to create meaningful and consistent assessment system.

Lecturers frequently employ multiple choices and essays for the test, particularly in theory course. A good test question must undergo item analysis before being used. This process is necessarily conducted to examine if the questions are composed based on principles of composing questions, and if they are reliable and valid.Djemari (2004:14) argues that the validity of measurement tool can be observed from its construction, i.e. measuring as planned. To find out if the questions composed by lecturers are good, an investigation into question items is necessarily conducted. The investigation is conducted on subject matter, construction, and language aspects. Subject matter aspect relates to science asked and level of thinking involved. Construction aspect relates to composing questions. Meanwhile, language aspect relates to the clarity of questions.

Scott (1993:146) posits that variations which are possibly conducted in developing written test are multiple-choice, sentence completion, listing, true-false, matching, essay, dan modified forms.

The next stage after question investigation is collecting empirical data through calibration (Djemari, 2012:182-184). A good test must conform validity and reliability aspects. Kusaeri & Suprananto (2012: 75- 82) argue that validity refers to appropriateness, meaningfulness, and usefulness. Meanwhile, reliability refers to consistency of a measurement. Based on the objectives of norm-referenced assessment, question items is not very difficult or very easy with difficulty index starts from 0.3 to 0.7 and can differentiate between smart students and not smart students with discrimination power index at least 0.3. Meanwhile, alternative answers must be selected by at least 5% of test takers (Djemari, 2008:143)

Through item analysis, it is possibly to investigate if composed questions qualify as good questions to measure students' learning achievement. Without item analysis, very easy questions or very difficult question are frequently found. The research result on item analysis of theory course in Hospitality and Fashion Education Study Programconducted by the researchers last year indicates that most of the final test questions are not validated by item analysis. Final test question are composed by not considering principles of composing a good test because it burdens lecturers. Besides, test is a form of comprehensive feedback which measure students' ability and skill. It is expected that test can reveal the students' real and undoubtful ability and skill. In general, this research aims to: 1) identify the quality of final test questions in written test; and 2) to identify the quality offinal tests questions of practice course.

2. Method

This is a survey research which aims to analyze the quality of final test questions for theory course and practice course composed by lecturers of the Faculty of Engineering. The survey consists of (1) item analysis for item difficulty level; (2) index of discrimination power; (3) item validation; and (4) reliability.

3. Procedure of Analysis

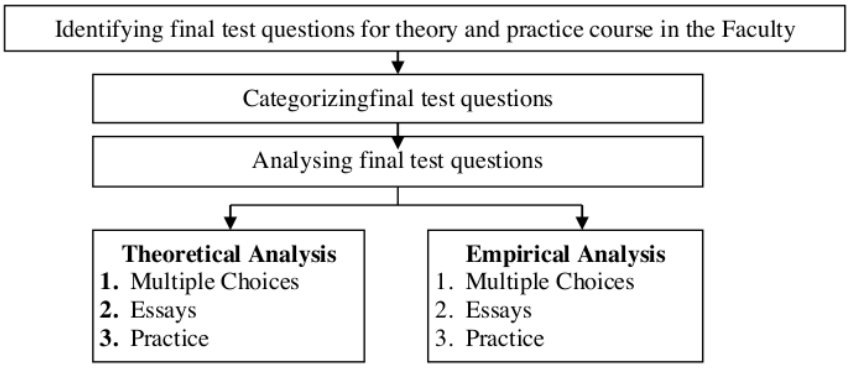

Figure 1. Research Procedure

This research was conducted in the Faculty of Engineering at Yogyakarta State University. The population of this research was final test questions for theory and practice course composed by the faculty's lecturers in even semester academic year 2017/2018. The research samples were then gained through purposive samplingtechnique; they were 17 multiple choice questions, 11 essay questions, and 11 practice test questions. The data were analysed by analysing final test questions theoretically and empirically. Theoretical analysis referred to item analysis while empirical analysis referred to difficulty levels, discrimination power, reliability, correlation between item score and total score, and quality of tricky question in written test of final test. Meanwhile, item analysis of practice test in the final test referred to item analysis and empirically analysed inter-rater score. Item analysis of written test employed Anates while item analysis of practice test employed Kappa.

4. Result

Table 1. The Summary of Empirical Quality of Multiple Choice Questions

Criteria Quality	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Very Poor (%)
Difficulty Level	0	0	70	30	0
Discrimination Power	0	6	64	30	0
Quality of Tricky Questions	0	41	0	59	0
Validity	0	0	86	12	0
Reliability	0	29	0	71	0

The analsysis result of multiple choice questions indicates that the quality of tricky questions still need to be considedered. Homogenous options are extremely expected to identify the students who have and have not studied.

Table 2. The Summary of Empirical Quality of Essay Questions

Quality	Very Good (%)	Good (%)	Fair (%)	Poor (%)	Very Poor (%)
Difficulty Level	0	0	70	30	0
Discrimination Power	0	18	54	28	0
Validity	0	0	64	36	0
Reliability	0	27	0	73	0

The quality of essay questions is relatively fair.However,the lecturers must observe them again because the result of item score correlation and total items shows that some questions are poor. Therefore, improvement is required. Theoretically, practice questions composed have reflected good questions. Empirically, items of practice course are assessed by three raters, and the result is presented in table 11.

Table 3. Summary of Empirical Quality of Practice Questions

Number of questions	Rater 1vs 2	Rater 1 vs 3	Rater 2 vs 3	Mean
1	0.80	0.63	0.70	0.71
2	0.83	0.65	0.60	0.68
3	0.64	0.68	0.70	0.67
4	0.80	0.80	1.00	0.87
5	0.70	0.83	0.75	0.83
6	0.80	0.70	0.62	0.71
7	0.60	0.60	0.78	0.66
8	0.80	0.80	0.80	0.80
9	0.70	0.80	0.80	0.76
10	0.80	0.85	0.80	0.82
11	0.68	0.70	0.75	0.71
			Mean Kappa	0.74

The analysis result of quality of practice questions from the three raters indicates the existence of consistency among three raters in analysing the quality of practice questions. The reliability index is > 0.7, and it supports the existence of assessment among raters. Furthermore, it indicates that the practice questions of final test are acceptably applied to assess the students' skills.

5. Conclusion

Based on the research results, it is concluded that: (1) The difficulty level of multiple choice questions of the final test is considered fair. It indicates that the questions are not really difficult and not really easy. (2) The discrimination level of multiple choice questions and essay questions considered fair. It indicates that the questions are able to distinguish students with high ability from students with low ability. (3) The effectiveness of tricky questions for multiple choice questions is considered poor. It indicates that the available options are not selected by all students. (4) The validity of multiple choice questions and essays questions is considered fair. It indicates that the items must be eliminated or revised. (5) The reliability of multiple choice questions and essay questions is poor. It indicates that the questions do not have good consistency. (6) The validity of practice questions is considered as very good. (7) The reliability of final test questions for practice course is good with the mean ofreliability index is > 0.7.

6. References

[1] Allen, M. J & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publising Company
[2] Berk, R.A. (1986). *Performance assessment*. Baltimore: The John Hopkins University Press
[3] Bloom B. S. (1956). *Taxonomy of educational objectives, Handbook I: The cognitive domain*. New York: David McKay Co, Inc.

[4] Brennan, R,L. (2006). *Educational measurement*. Westport: Praeger
[5] Bresciani, MJ, et al. (2009). Examining design and inter-rater reliability of a rubric measuring research quality across multiple disciplines. *Practical Assessment, Research & Evaluation*, Vol. 14, No 12
[6] Djemari Mardapi. (2004). Pengembangan sistem penilaian berbasis kompetensi. *Proceding*: Rekayasa sistem penilaian dalam rangka meningkatkan kualitas pendidikan. Yogyakarta: HEPI
[7] Djemari Mardapi. (2008). Teknik Penyusunan Instrumen Tes dan Nontes. Yogyakarta: Metra Cendekia
[8] Djemari Mardapi. (2012). Pengukuran Penilaian & Evaluasi Pendidikan. Yogyakarta: Nuha Medika
[9] Scott, J. L. (1993). *Improving vocational curriculum: Cognitif achievement evaluation*. Georgia: The Goodheart-Willcox Company, Inc
[10] Smith, M.V. (2007). Inter-rater reliability of flight school instructors: a foundational study. Diambil pada tanggal 20 September 2010 dari: *http://www.public.asu.edu/~mvsmith/IRRinAviation.pdf*.

C9 Quality analysis of final test in the faculty engineering of Yogyakarta State University

ORIGINALITY REPORT

12%	10%	8%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Program Pascasarjana Universitas Negeri Yogyakarta Student Paper	7%
2	kyutech.repo.nii.ac.jp Internet Source	3%
3	winarno.staff.iainsalatiga.ac.id Internet Source	1%
4	www.ijiet.org Internet Source	1%
5	media.neliti.com Internet Source	1%

Exclude quotes	Off	Exclude matches	< 1%
Exclude bibliography	On		

C9 Quality analysis of final test in the faculty engineering of Yogyakarta State University

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6