Abstract of 2017 AAOU Conference paper

The effect of NFD toward reliability and level of difficulty of multiple-choice question at Universitas Terbuka, Indonesia

Diki Diki, Eko Yuliastuti, Suratni Suratni

Universitas Terbuka, Indonesia

Parameters of quality of multiple-choice question (MCQ) are reliabilty, difficulty index (p), and distracter effectiveness (DE). DE reduces quality of test items. Meanwhile, reliability of each test item is measured by rpbis. Previous study revealed that an increase of 10 % in Non functioning distracter (NFD) would decrease the test reliability by 1. This study aims at measuring the effect of DE toward reliability of multiple-choice question (MCQ) test at Universitas Terbuka (UT), Indonesia. The sample is MCQ test at faculty of mathematics and natural sciences, Universitas Terbuka, Indonesia (Indonesia Open University). The researchers are conducting item analysis to gain p, rpbis, and NFD index. The measurement of relationship between p, DE, and NFD index is conducted through regression analysis. The result showed that DE affects significantly rpbis. However, DE does not always affect level of difficulty. Therefore, UT should consider NFD in the test development.

Key words: non-functioning distracter, multiple-choice question, distracter efficiency, reliabilty, level of difficulty.

**Introduction**

Universitas Terbuka (UT-Indonesia Open University) is the primary distance learning higher education institution in Indonesia. It was established in 1984. There are four faculties, which are Faculty of Law, Political and Social Sciences, Faculty of Economics, Faculty of Education and Teacher Training, and Faculty of Mathematics and Natural Sciences.

One core activity of UT is examination. Most examination methods are multiple-choice question (MCQ). In one examination in 2014, there were 37,247 students taking exams at UT. There were 1,627,382 test scores that were processed at the university (Adnan, 2015).

**Literatur study**

Multiple choice question (MCQ) consists of a stem and options. One options is the answer, while other options are distracters. If a distracter is chosen by one or more examinees, it is called functioning distracter. If a distracter is not chosen by less than 5% of examinees or less, it is called non-functioning distracter (Abdulghani et al, 2014; Deepak, et al, 2015).

Quality of MCQ consists of psychometric parameters. The parameters are reliabilty, difficulty index (p), and discrimination index (DI) (Abdulghani etl al, 2014; Rahma, 2017).

Non-functioning distracters is a distracter that is not chosen by less than 5% of the student. Previous studies showed that a set of MCQ may include one or more NFD. Rahmaetl. al (2017) found that 38 % of items of a set of a test has one NFD, 25% has two NFD, 33% has three NFD, and 2% has four NFD. On the other hand, Namdeo& Rout (2016) reported 16% of items has three NFD.

NFD influences level of difficulty of the test items. Abdulghani et al (2014) reported that NFD increases easiness of the item. They investigated the correlation between NFD and difficulty indices.

Similar results were also reported by Deepak et al (2015). They found that items with NFD have lower reliability. Previous study revealed that correlation between consistency and NFD is that an increase of 10 % in NFD would decrease the test reliability by 1 % (Deepak et. al, 2009).However, the study of Deepak et al used five-options MCQ, instead of four-options MCQ used at UT.

According to Namdeo& Rout (2016), a good quality MCQ should have no NFD (DE was 100%), in addition to good P (30% - 70%), and high DI (> 0,25). However, Namdeo and Rout did not study the effect of NFD toward difficulty, and reliability.

In addition, non-functioning distractor reduces quality of test items.Hingorjo& Jaleel (2013) found that having NFD did not reduce discriminatory indices.They found that items with one NFD has better discriminatory power (DI) than items with all four functioning distracters. The best discriminating items were those with three NFD.

However, the study of Hingorjo and Jaleel did not observe the effect of NFD toward reliability, DI, and p. They did a cross sectional study of MCQ.

Item analysis is a process of evaluating students’ responses to the test material by collecting and summarizing the responses. The most important result of an item analysis is its validity and reliability (Abdulghani et. al, 2015). For example, item analysis produces difficulty index, discriminatory index, and reliability index.

This study is to analyze the effect of NFD number in each items toward P, DI, and rpbis. Previous studies, such as Abdulghaniet. al (2015) did a study of NFD regarding its effect toward P, DI and reliability. However, the MCQ has five-option, instead of four that is used in UT. This study also tried to validate the study of Hingorjo& Jaleel (2013) who found that NFD did not affect P, DI, and reliability. As opposed to Hingorjo& Jaleel (2013) who used cross sectional method, this study used regression analysis.

**Rationale for the study**

The study about NFD may contribute to the improvement of the MCQ quality

No studies regarding NFD at Universitas Terbuka

**Research question**

1. To what extend do NFD affects pand reliability
2. What is the rate of NFD in the MCQ examination of PANG4312 course

**Aims of the study**

The study is to measure the effect of NFD toward test quality. Therefore, the researcher did a statistical analysis to find out the effect of NFD number toward Pand rpbis.

**Methodology**

The study did statistical analysis to find a relationship between NFD and psychometric indices of MCQ conducted at UT. The authors conducted regression analysis to find out if the NFD affect reliabilityand level of difficulty (Deepak, 2015).

The hypothesis of this study about the effect of DE toward item reliability was :

H0 = DE does not affect item reliability

H1 = DE affect item reliability

Mewanwhile, the hypothesis of this study about the effect of DE toward item reliability was :

H0 = DE does not affect item difficulty level

H1 = DE affect item difficulty level

We also measured mean, standard of deviation, and standard error of measurement.Effectiveness of distractors (DE) is measured by the number of NFD in each MCQ. If the question has no NFD, the DE is 100%. If a question has one NFD, the DE is 66%. If the question has two NFD, the DE is 33%. If the question has three NFD, the efficiency is 0% (Namdeo et. al, 2015).

Test reliability is measured by rpbis or the point biserial correlation, is a correlation between item and total test score. The value falls between -1 to 1. The positive value means students with higher total score are more likely to answer the item correctly. The negative value means students with higher score are more likely to answer the item incorrectly (Deepak et. al, 2015).

Difficulty index (P) is based on the number of correct answer out of total answer. This is a ratio of students who answer the item correctly to the total number of students taking the exam. Items with P < 0,02 is very difficult. Items with P between 0,21 and 0,69 is moderately difficult. Items with P between 0,7 and 0,9 is easy, while items with P > 0,9 is very easy (Abdulghani et. al, 2014; Haladyna & Downing, 1989).

There were six samples of this study. Each sample was an examination set. Those samples were BIOL4110 2014.2., BIOL4110 2015.2, PANG4214 2015.2, PANG4214 2015.2, BIOL4417 20151, and BIOL4223 2015.2. Each examination set consists of 45 item. Total students participating in those test were 1491 students. The researcher gained the examination result from the Examination Center of UT.

**Result and discussion**

In the sample of BIOL4110 2014.2, there were 458 students. KR20 was 0,609. There were 11 items with 1 NFD and 1 item with 2 NFD.

The regression analysis for the effect of DE toward reliability showed that F was 3.336, R2 was 20,6%, while p was 0,02 which was < 0,05. The regression model statistically significantly predicts the outcome variable.DE positively affects reliability. This result is in line with the result of Deepak et. al (2015) that if the DE is higher, the reliability will be higher.

The regression result for the effect of DE toward difficulty level showed that R2 was 27,3%, F was 16.631, and sig t was 0,00. The result shows that DE positively affects level of difficulty. This result is in line with the result of Abdulghani et al (2014) that an increase in DE will improve difficulty index.

Int the sample of BIOL4110 2015.2, there were 524 students with KR 20 was 0,60. The regression analysis for the effect of DE toward reliability showed that F was 5.336, R2 was 11 %, while p was 0,03 which was < 0,05. The regression model statistically significantly predicts the outcome variable.DE positively affects reliability. This result is in line with the result of Deepak et. al (2015) that if the DE is higher, the reliability will be higher.

However, DE did not affect level of difficulty. The sig t was 0.14, which was greater than 0,05.

In the sample of PANG4214 2015 2, there were 128 students and KR20 was 0,69. There were 8 items with one NFD each. The regression result for the effect of DE toward reliability shows that R2 was 9,3%, F was 4.419, and sig t was 0,04. The result shows that DE positively affects level of reliability. This result is in line with the result of Deepak et. al (2015) that if the DE is higher, the reliability will be higher.

The regression result for the effect of DE toward difficulty level showed that R2 was 10,5%, F was 5,035, and sig t was 0,03. The result shows that DE positively affects level of difficulty. This result is in line with the result of Abdulghani et al (2014) that an increase in DE will improve difficulty index.

In the sample of BIOL4417 20151, there were 250 students and KR20 was 0,69. The regression result for the effect of DE toward reliability shows that F was 6.394, R2 was 12,9%, and sig t was 0.015. The result shows that DE positively affects level of reliabilty.

The regression result for the effect of DE toward difficulty level showed that R2 was 1,2%, F was .50 and sig t was 0,48. The result shows that DE does not affect difficulty level. This result is not in line with the result of Abdulghani et al (2014) that an increase in DE will improve difficulty index.

In the BIOL4223 2015.2 the NFD did not affect reliability, since sig t was 0,69. However, NFD affect difficulty level, with sig t of 0,00. KR 20 was 0,35 and the student number was 131 and KR20 was 0,61.

**Conclusion**

The result of this study supports the finding of Deepak et. al (2015) that incidence of NFD affects reliability of test item. In addition, this study used regression analysis, as opposed to Deepak et. al who used correlation.

On the other hand, NFD does not always affect difficulty level of the item. Previous study of Abdulghani (2013) disclosed otherwise. For example, NFD affected difficulty level in BIOL4110 2014.2. However, the result in BIOL4110 2015.2 showed otherwise.

This study showed that the test development at UT should take into account the distracter effectiveness. Test developers should avoid making non-functioning distracters. The NFD reduces item reliability. The incidence of NFC also make the item easier.

**References:**

Abdulghani, H. M., Ahmad, F., Ponnamperuma, G. G., Khalil, M. S., & Aldrees, A. (2014). The relationship between non-functioning distractors and item difficulty of multiple choice questions: A descriptive analysis. *Journal of Health Specialties*, *2*(4), 148.

Adnan, I. (2015). Student Assessment at Universitas Terbuka Indonesia.

Deepak, K. K., Al-Umran, K. U., AI-Sheikh, M. H., Dkoli, B. V., & Al-Rubaish, A. (2015). Psychometrics of Multiple Choice Questions with Non-Functioning Distracters: Implications to Medical Education. *Indian journal of physiology and pharmacology*, *59*(4), 428.

Haladyna, T. M. & Downing, S. M. (1989). A taxonomy of multiple choice-item writing rules.

Hingorjo, M. R., & Jaleel, F. (2012). Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *JPMA-Journal of the Pakistan Medical Association*, *62*(2), 142.

Rahma, N. A., Shamad, M. M., Idris, M. E., Elfaki, O. A., Elfakey, W. E., & Salih, K. M. (2017). Comparison in the quality of distractors in three and four options type of multiple choice questions. *Advances in Medical Education and Practice*, *8*, 287.