**Artikel Penelitian**

# VALIDATION OF LEARNING EVALUATION QUESTIONNAIRE
# IN FACULTY OF MEDICINE PATTIMURA UNIVERSITY

Nerissa Alviana Sutantie[1], Farah Christina Noya[1], Rif'ah Zafarani Soumena[1]

[1]Medical Faculty of Pattimura University

*Corresponding author* e-mail **:** nsutantie@gmail.com

**Abstrak**

**Introduction.** Efforts to guarantee the quality of medical doctor graduates are major concern in the implementation of problem-based learning in medical education today. Therefore, continuous evaluation in the medical education system is necessary. However, to obtain constructive evaluation results, the instrument that is used must be valid and reliable. **Aim** This study aimed to validate the learning evaluation questionnaire used at the Faculty of Medicine Pattimura University, so the valid and reliable instrument can be produced for further evaluation. **Methods** This study used correlative analytic method; with study subject are 102 students that were selected randomly through stratified random sampling technique and had been adjusted to restriction criteria. Pearson bivariate correlation test was performed to obtain the validity coefficient and the reliability coefficient by Alpha method. **Result** The results of the initial analysis and the second analysis after modification show that some items have low validity (0.311 and 0.256) and even invalid (- 0.03; < r table 0.1946). However, after item reconstruction, the average validity coefficient raise to >0.5 (moderate category) and >0.9 (very high category) with the overall questionnaire reliability coefficient is >0.9 in two tests (test-retest). **Conclution.** Thus, it can be concluded that the learning evaluation questionnaire used at the Faculty of Medicine Pattimura University is valid and reliable.

*Keywords* **:** Validity, Reliability, Learning Evaluation.

## Introduction

The effort to guarantee the quality of doctor graduates is a major concern in the implementation of medical education today.[1,2,3] Therefore, continuous evaluation in the medical education system is needed.[1,2] Related to accreditation, data of KKI Year 2016[5] shows Indonesia still dominated by FK with accreditation of C (37 FK from total 75 FK) while FK outcome in relation to certification also not satisfy to see number of passing of some FK less than 50%. Graduate accreditation and certification is an example of an evaluation from an external FK. In addition to external evaluations, FKs should also conduct an internal evaluation of the institution.

Internal evaluation is mostly done using the scale of students satisfaction through questionnaires.[4,5,6] Nevertheless, until now the assessment of the effectiveness of learning and teaching has not been effective.[12] One is influenced by many instruments developed of the faculty (homemade) so that there is doubtful validity and reliability.[5,7,8]

The above findings are consistent with several systematic reviews that indicate the low evidence of validity in an evaluation method in medical education by Davis et al[9] that 9 out of 17 studies used pretested or prevalidated measures. A review of evaluation instruments for medical professionalism by Veloski et al[10] found only 11% of studies featuring strong evidence with exposure to the results of calculations of content validity, and internal consistency reliability. In addition, peer assessment-related findings indicate that until

now there has not been a universal instrument and depends only on the assessment of faculty as a gold standard in developing the quality of institutional education.

Regarding the importance of an evaluation, Faculty of Medicine Pattimura University (Unpatti) has also conducted a gradual evaluation in its institution through completing questionnaires by students on an ongoing basis at the end of each semester. The learning evaluation questionnaire at FK Unpatti is a questionnaire adapted by MEU from the learning evaluation questionnaire of Maastrich University, Netherland and has been used from 2012-2016. The results of these evaluations have actually been helpful even though they have not been validated since the beginning of their use. Thus, to be able to apply the evaluation results on a larger scale such the improvement of curriculum or accreditation would need to be validated first.

**Methods**

This study is a correlative analytic with validity test using Pearson bivariate correlation test and reliability test through Merode Alpha with test-retest technique. The population in this study are 102 students of Faculty of Medicine Pattimura University academic year 2016/2017 (semester 2, 4, and 6) by stratified random sampling technique and according to restriction criterion.

Data in this study is primary data obtained by filling out the questionnaire of learning evaluation made in the form of web based questionnaire, and filling up 3 times at

different time. The first questionnaire was filled out by completing the initial questionnaire from of MEU FK Unpatti conducted on 20 June 2017, followed by the second questionnaire, modified and reconstructed questionnaire (test) on July 14, 2017 and third retest (questionnaire 2) on July 24, 2017.

The data obtained from the questionnaire of learning evaluation at FK Unpatti was processed using Microsoft Office Excel 2007 and Software Statistic Packages for Social Sciences (SPSS) for Windows version 24.0. To obtain the coefficient of validity used Pearson bivariate correlation test while reliability coefficient obtained through the Alpha method.

The validity of a statement item is determined by comparing the correlation coefficients that have been obtained with the r table. Researcher use r table value according to degree of freedom equal to 100 (dk = n-2) with significance 0,05 was 0,1946 whereas item statement stated reliable if value of Alpha> r table (at 5% significance level) [11,12] Next, the coefficients of validity and reliability are matched by the category of validity and reliability according to Guilford.[13]

**Results**

*Validity*

Validity test is done for each statement of indikator 1-7 in two stages:

1. Initial validity test and modification

The researcher tested the initial validity using a learning evaluation questionnaire from MEU that was created in the form of a web-

http://ojs3.unpatti.ac.id/index.php/moluccamed

based questionnaire. The point of the questionnaire statement is still in its original form and structure only in the part of the "batch" changed to "semester" to match the research criteria. (see table 4.1 – 4.7)

2. Test of Reconstruction Validity

The results of the reconstruction validity test obtained generally increased the coefficient of validity but there is also a fixed value. (see table 4.8- 4.11)

Specification of the validity test result:

Learning evaluation questionnaire generated in this research outline has the same question /statement item with the initial questionnaire only there are some changes that is:

1. The modified questions are listed in indicator 1 (number 7, 8) and indicator 2 (number 21 and 22).

2. For statements reconstructed in indicator 1 (number 6), indicator 2 (number 19, 20, 21, 22 and 23), indicator 5 (number 48) and indicator 6 (number 60, 61, 62). In addition to modification and reconstruction, there are questions that are shaped from closed-ended questions to open-ended questions, namely questions 22 and 23 on indicator 2.

*Reliability*

Reliability test in this study was conducted after the questionnaire was reconstructed and declared at minimum moderate validity

category. Test reliability is done twice (test-retest) in different time (set time interval is 10 days).

1. Reliability test I (test)

In accordance with table 4.12 looks reliability test result of this questionnaire (indicator 1-7) in general very good. Measurement of reliability as a whole is also obtained coefficient Alpha 0.985 so that the questionnaire declared reliable with very high category.

2. Reliability test II (*test*)

The results of the repeatability test contained in Table 4.13 (appendix 2) generally show that the Alpha coefficient value is consistent both the reliability value per indicator and the total reliability (0.984) although there is a decrease in value but not significant.

**Discussion**

*Questionnaire Validity*

Based on the results of the above research, after modification and reconstruction, the validity score for the statement items in the questionnaire is generally good, including high category (23.5%) and very high (70.5%) and medium (6%). Although previously there are items of statements that have low validity even invalid.

**Table 4.1 Validity of question indicator 1**

| Statement Item Number | Initial Validity | | | | After Modified | |
|---|---|---|---|---|---|---|
| | Validity Coefficient | r table | Result | Validity Category | Validity Coefficient | Validity Category |
| 1. | 0.489 | 0.1946 | Valid | Moderate | 0.650 | High |
| 2. | 0.559 | 0.1946 | Valid | Moderate | 0.676 | High |
| 3. | 0.617 | 0.1946 | Valid | High | 0.734 | High |
| 4. | 0.536 | 0.1946 | Valid | Moderate | 0.660 | High |
| 5. | 0.627 | 0.1946 | Valid | High | 0.753 | High |
| 6. *) | 0.280 | 0.1946 | Valid | Low | 0.311 | Low |
| 7. *) | 0.674 | 0.1946 | Valid | High | 0.564 | Moderate |
| 8. *) | 0.674 | 0.1946 | Valid | High | 0.564 | Moderate |

*) Statement/question modified

**Table 4.2 Validity of question indicator 2**

| Statement Item Number | Initial Validity | | | |
|---|---|---|---|---|
| | Validity Coefficient | r table | Result | Validity Category |
| 10. | 0.744 | 0.1946 | Valid | High |
| 11. | 0.728 | 0.1946 | Valid | High |
| 12. | 0.730 | 0.1946 | Valid | High |
| 13. | 0.847 | 0.1946 | Valid | Very High |
| 14. | 0.829 | 0.1946 | Valid | Very High |
| 15. | 0.892 | 0.1946 | Valid | Very High |
| 16. | 0.858 | 0.1946 | Valid | Very High |
| 17. | 0.879 | 0.1946 | Valid | Very High |
| 18. | 0.816 | 0.1946 | Valid | Very High |
| 19. | 0.844 | 0.1946 | Valid | Very High |
| 20. | 0.871 | 0.1946 | Valid | Very High |
| 21. *) | 0.668 | 0.1946 | Valid | High |
| 22. *) | - 0.037 | 0.1946 | Invalid | Invalid |
| 23. *) | 0.272 | 0.1946 | Valid | Low |

| Statement Item Number | After Modified | |
|---|---|---|
| | Validity Coefficient | Validity Category |
| 10. | 0.769 | High |
| 11. | 0.747 | High |
| 12. | 0.759 | High |
| 13. | 0.865 | Very High |
| 14. | 0.848 | Very High |
| 15. | 0.910 | Very High |
| 16. | 0.867 | Very High |
| 17. | 0.893 | Very High |
| 18. | 0.831 | Very High |
| 19. | 0.854 | Very High |
| 20. | 0.883 | Very High |
| 21. *) | 0.582 | Moderate |
| 22. *) | - 0.03 | Invalid |
| 23. *) | 0.256 | Low |

*) Statement/question modified

**Table  4.3 Validity of question indicator 3**

| Statement Item Number | Validity Coefficient | r table | Result | Validity Category |
|---|---|---|---|---|
| 25. | 0.718 | 0.1946 | Valid | High |
| 26. | 0.853 | 0.1946 | Valid | Very High |
| 27. | 0.826 | 0.1946 | Valid | Very High |
| 28. | 0.733 | 0.1946 | Valid | Very High |
| 29. | 0.886 | 0.1946 | Valid | Very High |
| 30. | 0.846 | 0.1946 | Valid | Very High |
| 31. | 0.808 | 0.1946 | Valid | Very High |
| 32. | 0.860 | 0.1946 | Valid | Very High |

**Table  4.4 Validity of question indicator 4**

| Statement Item Number | Validity Coefficient | r table | Result | Validity Category |
|---|---|---|---|---|
| 34. | 0.873 | 0.1946 | Valid | Very High |
| 35. | 0.787 | 0.1946 | Valid | Very High |
| 36. | 0.850 | 0.1946 | Valid | Very High |
| 37. | 0.801 | 0.1946 | Valid | Very High |
| 38. | 0.848 | 0.1946 | Valid | Very High |
| 39. | 0.867 | 0.1946 | Valid | Very High |
| 40. | 0.869 | 0.1946 | Valid | Very High |
| 41. | 0.764 | 0.1946 | Valid | High |
| 42. | 0.731 | 0.1946 | Valid | High |

**Table  4.5  Validity of question indicator  5**

| Statement Item Number | Validity Coefficient | r table | Result | Validity Category |
|---|---|---|---|---|
| 44. | 0.815 | 0.1946 | Valid | Very High |
| 45. | 0.822 | 0.1946 | Valid | Very High |
| 46. | 0.607 | 0.1946 | Valid | Very High |
| 47. | 0.895 | 0.1946 | Valid | Very High |
| 48. | 0.580 | 0.1946 | Valid | Moderate |
| 49. | 0.786 | 0.1946 | Valid | High |
| 50. | 0.860 | 0.1946 | Valid | Very High |
| 51. | 0.891 | 0.1946 | Valid | Very High |
| 52. | 0.650 | 0.1946 | Valid | High |
| 53. | 0.693 | 0.1946 | Valid | High |

**Table 4.6 Validity of question indicator 6**

| Statement Item Number | Validity Coefficient | r table | Result | Validity Category |
|---|---|---|---|---|
| 55. | 0.890 | 0.1946 | Valid | Very High |
| 56. | 0.891 | 0.1946 | Valid | Very High |
| 57. | 0.895 | 0.1946 | Valid | Very High |
| 58. | 0.875 | 0.1946 | Valid | Very High |
| 59. | 0.885 | 0.1946 | Valid | Very High |
| 60. | 0.579 | 0.1946 | Valid | High |
| 61. | 0.338 | 0.1946 | Valid | Low |
| 62. | 0.635 | 0.1946 | Valid | High |
| 63. | 0.718 | 0.1946 | Valid | High |
| 64. | 0.759 | 0.1946 | Valid | High |

**Table 4.7 Validity of question indicator 7**

| Statement Item Number | Validity Coefficient | r table | Result | Validity Category |
|---|---|---|---|---|
| 66. | 0.892 | 0.1946 | Valid | Very High |
| 67. | 0.906 | 0.1946 | Valid | Very High |
| 68. | 0.875 | 0.1946 | Valid | Very High |
| 69. | 0.921 | 0.1946 | Valid | Very High |
| 70. | 0.938 | 0.1946 | Valid | Very High |
| 71. | 0.865 | 0.1946 | Valid | Very High |
| 72. | 0.871 | 0.1946 | Valid | Very High |
| 73. | 0.789 | 0.1946 | Valid | Very High |
| 74. | 0.871 | 0.1946 | Valid | VeryHigh |
| 75. | 0.744 | 0.1946 | Valid | High |

**Table 4.8 Validity of reconstructed statement item indicator 1**

| Statement Item Number | Validity Coefficient | r table | Result | Validity Category |
|---|---|---|---|---|
| 1. | 0.767 | 0.1946 | Valid | High |
| 2. | 0.771 | 0.1946 | Valid | High |
| 3. | 0.738 | 0.1946 | Valid | High |
| 4. | 0.651 | 0.1946 | Valid | High |
| 5. | 0.764 | 0.1946 | Valid | High |
| 6.*) | 0.537 | 0.1946 | Valid | Moderate |
| 7. | 0.576 | 0.1946 | Valid | Moderate |
| 8. | 0.569 | 0.1946 | Valid | Moderate |

*)Statement/question reconstructed

**Table 4.9 Validity of reconstructed statement item indicator 2**

| Statement Item Number | Validity Coefficient | r table | Result | Validity Category |
|---|---|---|---|---|
| 10. | 0.808 | 0.1946 | Valid | Very High |
| 11. | 0.876 | 0.1946 | Valid | Very High |
| 12. | 0.878 | 0.1946 | Valid | Very High |
| 13. | 0.798 | 0.1946 | Valid | Very High |
| 14. | 0.845 | 0.1946 | Valid | Very High |
| 15. | 0.858 | 0.1946 | Valid | Very High |
| 16. | 0.834 | 0.1946 | Valid | Very High |
| 17. | 0.836 | 0.1946 | Valid | Very High |
| 18. | 0.860 | 0.1946 | Valid | Very High |
| 19. *) | 0.866 | 0.1946 | Valid | Very High |
| 20. *) | 0.861 | 0.1946 | Valid | Very High |
| 21. **) | 0.734 | 0.1946 | Valid | High |
| 22. **) | 0.373 | 0.1946 | Valid | Low |
| 23. **) | 0.335 | 0.1946 | Valid | Low |
| **Without question number 22 dan 23** | | | | |
| **Number** | **Validity Coefficient** | | **Validity Category** | |
| 10. | 0.848 | | Very High | |
| 11. | 0.895 | | Very High | |
| 12. | 0.926 | | Very High | |
| 13. | 0.857 | | Very High | |
| 14. | 0.883 | | Very High | |
| 15. | 0.914 | | Very High | |
| 16. | 0.900 | | Very High | |
| 17. | 0.888 | | Very High | |
| 18. | 0.843 | | Very High | |
| 19. *) | 0.933 | | Very High | |
| 20. *) | 0.914 | | Very High | |
| 21. **) | 0.805 | | Very High | |

*) Statement/question reconstructed

**) Statement/question modified and reconstructed

**Table 4.10 Validity of reconstructed statement item indicator 5**

| Statement Item Number | Validity Coefficient | r table | Result | Validity Category |
|---|---|---|---|---|
| 44. | 0.909 | 0.1946 | Valid | Very High |
| 45. | 0.838 | 0.1946 | Valid | Very High |
| 46. | 0.863 | 0.1946 | Valid | Very High |
| 47. | 0.876 | 0.1946 | Valid | Very High |
| 48. *) | 0.855 | 0.1946 | Valid | Very High |
| 49. | 0.854 | 0.1946 | Valid | Very High |
| 50. | 0.861 | 0.1946 | Valid | Very High |
| 51. | 0.921 | 0.1946 | Valid | Very High |
| 52. | 0.841 | 0.1946 | Valid | Very High |
| 53. | 0.762 | 0.1946 | Valid | Very High |

*) Statement/question reconstructed

**Table 4.11 Validity of reconstructed statement item indicator 6**

| Statement Item Number | Validity Coefficient | r table | Result | Validity Category |
|---|---|---|---|---|
| 55. | 0.907 | 0.1946 | Valid | Very High |
| 56. | 0.880 | 0.1946 | Valid | Very High |
| 57. | 0.844 | 0.1946 | Valid | Very High |
| 58. | 0.844 | 0.1946 | Valid | Very High |
| 59. | 0.864 | 0.1946 | Valid | Very High |
| 60. *) | 0.849 | 0.1946 | Valid | Very High |
| 61. **) | 0.791 | 0.1946 | Valid | Very High |
| 62. *) | 0.587 | 0.1946 | Valid | Very High |
| 63. | 0.828 | 0.1946 | Valid | Very High |
| 64. | 0.694 | 0.1946 | Valid | High |
| 65. | 0.656 | 0.1946 | Valid | High |

*) Statement/question reconstructed

**) New Statement (separated from statement item number 60)

**Table 4.12 Reliability I (*Test*)**

| Indicator | Reliability Coefficient (α) | r table | Result | Reliability Category |
|---|---|---|---|---|
| 1. | 0.799 | 0.1946 | Reliable | Very High |
| 2. | 0.974 | 0.1946 | Reliable | Very High |
| 3. | 0.927 | 0.1946 | Reliable | Very High |
| 4. | 0.956 | 0.1946 | Reliable | Very High |
| 5. | 0.958 | 0.1946 | Reliable | Very High |
| 6. | 0.931 | 0.1946 | Reliable | Very High |
| 7. | 0.971 | 0.1946 | Reliable | Very High |
| Total | 0.985 | 0.1946 | Reliable | Very High |

**Tabel 4.13 Reliability II (Re-*Test*)**

| Indicator | Reliability Coefficient (α) | r table | Result | Reliability Category |
|---|---|---|---|---|
| 1. | 0.768 | 0.1946 | Reliable | Very High |
| 2. | 0.977 | 0.1946 | Reliable | Very High |
| 3. | 0.936 | 0.1946 | Reliable | Very High |
| 4. | 0.954 | 0.1946 | Reliable | Very High |
| 5. | 0.952 | 0.1946 | Reliable | Very High |
| 6. | 0.915 | 0.1946 | Reliable | Very High |
| 7. | 0.976 | 0.1946 | Reliable | Very High |
| Total | 0.984 | 0.1946 | Reliable | Very High |

To obtain the validity value as described, the statement item has gone through several stages of improvement namely the modification stage by simply changing the category of answer choice as well as the reconstruction phase by changing the sentence structure in the statement item on the results of the initial validity test and after modification, there are several statements that are reconstructed with the possible factors that influence [14,15] ie the

statement contains words with broad meaning /difficult understood, statement with possible answers vary/not homogeneous, statements have 2 intentions in 1 sentence / not specific, and the structure of writing that is less precise. In general, the researcher reconstructs some of the above statements by taking into account conditions such as: statements relevant to survey objectives, easy to ask, easy to answer, and easily obtained data.[15]

In addition to the above findings, the researcher also found a statement with high validity value but actually not applied well now is about the use of internet in FK for learning process. After analyzing the likelihood of the respondent's response is influenced by the condition prior to the transfer from the FK campus in PGSD which has provided complete internet access (applicable) and the possibility of providing FK campus internet facility in POKA which is more complete in the last few months.

*Questionnaire Reliability*

Based on the results of this study, the overall Alpha coefficient on the first measurement is 0.985 and the second measurement is 0.984. In accordance with the reliability coefficients set by Guilford,[13] Sukardi,[16] and the rule of thumb George and Mallery[17] the questionnaire has a very high reliability (> 0.9) and has met the recommended Alpha coefficient (> 0.7 or min> 0, 6).

In addition, when viewed from the method of reliability testing used reliability test results can be said to be satisfactory because it
56

has been through testing twice (test-retest) at different times. The choice of time interval in this study is 10 days referring to reference[18] which says that the time between testing 1 and 2 is generally a few days or weeks only.

Actually behind the high or low reliability of a questionnaire, there are several factors that affect. The researchers were aware of the factors mentioned by Djaali and Muljono[19] and Miller et al[20] on the instruments and self-respondents during the data retrieval although in the end the researchers did not find any low reliability values due to these factors.

*Relationship of questionnaire validity and reliability*

The result of the study shows that the learning evaluation questionnaire at FK Unpatti has validity and reliability value that is high questionnaire validity value, very high and minimal medium category supported by very high reliability value. This is in accordance with the theory by Arikunto[21] which states that a valid questionnaire is generally reliable. In addition, referring to the illustrations in Bolarinwa[22] explains that a valid and reliable questionnaire means the instrument is capable of achieving the stated measurement objectives.

**Conclusions**

Based on the results above, it can be concluded that the learning evaluation questionnaire at FK Unpatti has been feasible to be used with validity value after modification and reconstruction of the statement items including high and very high validity category and minimal medium/ moderate category while
http://ojs3.unpatti.ac.id/index.php/moluccamed

reliability in first or second test included category very high. Thus, the questionnaire produced can be used as a instrument for the next study in conducting an analysis of the evaluation results in FK Unpatti as well as an authentic evidence for the institution that had been validated on the evaluation questionnaire.

**Reference**

1. Stalmeijer R, Dolmans D, Van Berkel H, Wolfhagen I. *Quality assurance*. In: Van Berkel H, Scherpbier Hillen H, Van der Vleuten C, editors. *Lessons from Problem-based learning*. New York: Oxford University Press; 2010. p.157-158.

2. World Federation For Medical Education. *Basic medical education* WFME *global standards for quality improvement the 2015 revision*. WFME Office; 2015.

3. Konsil Kedokteran Indonesia. Standar pendidikan profesi dokter indonesia. Jakarta: Konsil Kedokteran Indonesia, 2012.

4. Schiekirka S, Reinhardt D, Heim S, Fabry G, Pukrop T,  Anders S et al. *Student perceptions of evaluation in undergraduate medical education: a qualitative study from one medical school*. BMC Medical Education. 2012;12:45.

5. Fleming P, Heath O, Goodridge A, Curran V. *Making medical student course evaluations meaningful: implementation of an intensive course review protocol*. BMC Med Educ. 2015;1:99.

6. Agarwal A, Wong S, Sarfaty S, Devaiah A, Hirsch. *Elective courses for medical students during the preclinical curriculum: a systematic review and evaluation*. Med Educ Online. 2015;20: 26615.

7. Colbert-Getz JM, Baumann S. *Changing medical students' perception of the evaluation culture: Is it possible?*. J Educ Eval Health Prof. 2016;13:8.

8. Annan SL, Tratnack S, Rubenstein C, Metzler-Sawin E, Hulton L. *An integrative review of student evaluations of teaching: implications for evaluation of nursing faculty*. J Prof Nurs. 2013 Sep-Oct; 29(5):10-24.

9. Davis DA, Mazmanian PE, Fordis M, VanHarrison R, Thorpe KE, Perrier L. *Accuracy of physician self-assessment compared with observed measures of competence: a systematic review*. JAMA. 2006;296:1094–1102.

10. Veloski JJ, Fields SK, Boex JR, Blank LL. *Measuring professionalism: a review of studies with instruments reported in the literature between 1982 and 2002*. Acad Med. 2005;80:366 –370.

11. Triton PB. SPSS 13.0 terapan: riset statistik parametrik. Jogjakarta: Andi Ofset; 2005.

12. Widiyanto J. *SPSS for windows.* Surakarta: Universitas Muhammadiyah Surakarta; 2012.

13. Guilford JP. *Fundamental statistics in psychology and education*. New York: Mc Graw-Hill Book Co. Inc ;1956.

14. Kazi MA, Khalid W. *Questionnaire designing and validation*. J Pak Med Assoc. May 2012; 62: (5).

15. Evergreen S, Gullickson A, Mann C, Welch W. *Handout for the webinar on developing & validating survey instruments*. Western Michigan University; 2011.

16. Sukardi. Evaluasi pendidikan dan prinsip operasionalnya. Jakarta: Bumi Aksara; 2008.

17. George D, Mallery. *SPSS for windows step by step: a simple guide and reference 11.0 Update*. Allyn and Bacon. Boston; 2003.

18. Test reliability. *Proffesional testing inc*; 2006.

19. Djaali H, Muljono P. Pengukuran dalam bidang pendidikan. Jakarta: PT Gramedia Widiasarana; 2008.

20. Miller MD, Lin RL, Gronlund NE. *Measurement and assessment in teaching tenth edition.* New Jersey: Pearson Education Inc; 2009.

21. Arikunto S. Dasar-dasar evaluasi pendidikan (edisi 2). Jakarta: Bumi Aksara; 2012.

22. Bolarinwa OA. *Principles and methods of validity and reliability testing of questionnaires used in social and health science researches*. Nigerian Postgraduate Medical Journal. 2015;(22):(4)