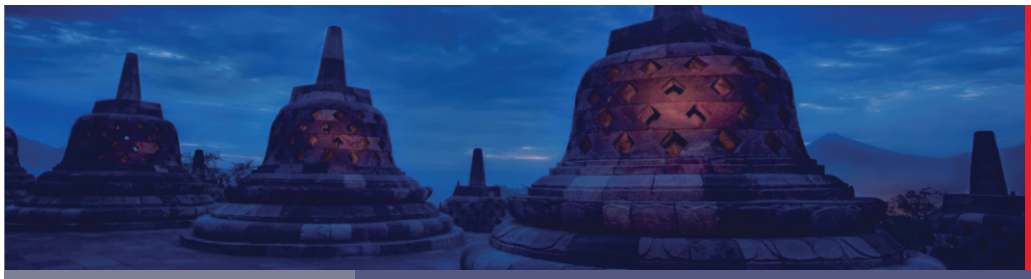

ISSN: 2407-1501

# PROCEEDING

# INTERNATIONAL CONFERENCE ON EDUCATIONAL RESEARCH AND EVALUATION (ICERE)

*"Assessment for Improving Students' Performance"*

**May 29 – 31 2016**
**Rectorate Hall and Graduate School**
**Yogyakarta State University**
**Indonesia**

**Proceeding**
International Conference on Educational Research and Evaluation (ICERE) 2016

**Publishing Institute**
Yogyakarta State University

**Director of Publication**
Prof. Djemari Mardapi, Ph.D.



**Address**
Yogyakarta State University
ISSN: 2407-1501

## Table of Contents

# ANALYSIS OF INTER-RATER CONSISTENCY IN ASSESSMENT FINAL PROJECT FASHION STUDY PROGRAM

Emy Budiastuti
Fashion Study Program (Faculty of Engineering, YSU)
E-mail: emy.budiastuti@yahoo.com

***Abstract***— This study aims at analyzing score to obtain *Inter-Rater* consistency in the Final Project assessment year 2015 of the Clothing Design Education Study Program based on internal and external assessments. This research is a survey research. The sample of this research is the sixth semester students of Undergraduate Program. The rater as the external experts (academics, practitioners, and association), while the internal experts are the advisors. The coefficient of consistency based on the ICC analysis < 0.50 and Alpha < 0,70. The ICC criteria is 0,50 at minimum, while Alpha coefficient at least 0.70. The magnitude of ICC coefficient is strongly influenced by the observed value range (variability – variance). Meanwhile, the magnitude of Alpha coefficient is strongly influenced by the number of samples.

***Keywords:*** *consistency analysis, inter-rater, final project*

## I. INTRODUCTION

Final Project course is a compulsory course for the students of the Faculty of Engineering of Yogyakarta State University, especially the students of Clothing Design Education study program which they are obliged to take and pass. According to its name, the students are demanded to make a fashion project that is going to display in a fashion show with the theme prescribed. Besides, the students are requested to make a report which later will be presented in the Final Project examination.

There are two stages of the Final Project implementation, i.e. fashion show and oral test. The assessment technique is also adjusted with the two stages. In the fashion show, clothing products made by the students are shown by models, but they must be assessed previously. The assessment of the show is done by the external parties (academics, professional association and practitioners). Meanwhile, the products are assessed by the internal parties, i.e. advisors, and later the results will be combined by the results of oral test.

The assessment of clothing product is started by internal and external parties. They are demanded to give actual scores. *Performance Assessment* of the final project is an assessment that demands the test participant or the students to demonstrate knowledge and skills according to the expected criteria and to be able apply them actually. The students' knowledge and skills can be found out through scoring. Scoring is something that should be concerned, because through scoring, the students' actual abilities will be revealed. The problems which are frequently found in arranging and using skill test lying on validity, reliability and *fairness*.

Performance assessment is an assessment which is done by observing the students' activity while they are doing something. The performance assessment is appropriate to assess competency achievement which demands the students to conduct certain duty, i.e. creating fashion products. This kind of assessment is considered more authentic than written test because the object of assessment reflects the actual ability of the students.

Rater in the Final Project assessment plays important role. *Random error* from the rater will influence the students' score difference comprehensively. There are three sources of mistakes in skill assessment scoring, i.e.: 1) instrument problems; 2) procedural problems; and 3) biased scoring problems. To get result or actual score of the students, the consistency of *inter-rater* is needed much in designing appropriate scoring rubric, selecting and training for *raters*, and *rechecking rater performance*. Based on research conducted by Wainer and Thissen (1993), one of sustainable issues in performance-based assessment is tidak adanya keandalan skala karena

*rater.* Robert (1981) explained that in order to minimize the measuring mistakes on the performance test, it is suggested to use some raters in assessing the students' behaviors and actually consistency of raters is very determining.

Based on Final Project assessment in the fashion show held previously, the results raise dissatisfaction of the students and advisors. This is because the right of decision is absolutely belongs to the expert. After being observed, there is less understanding or agreement in doing assessment. Therefore, in 2015, the researcher will make efforts to analyze the assessment consistency done by the expert so the dissatisfaction of the students and the advisors will not arise. Through this research, it is expected that there is consistency of *inter-rater* in the assessment of Final Project as expected.

The purpose of this research are: 1) to analyze score in order to create consistency of *Inter-Rater* in the Final Project assessment in 2015 of the Fashion Design Education Study Program based on the external assessment, 2) to analyze score to get *Inter-Rater* consistency in the Final Project year 2015 of the students of Fashion Design Education Study Program based on internal assessment.

The assessment is very important in learning. It must be based on objective, reliable, and trustworthy principles, it should motivate the students to keep on learning. As explained by Kusaeri (2012:16), assessment is a systematic procedure and involved activity of collecting, analyzing, as well as interpreting information that can be used to make conclusion regarding someone's characteristics to determine how extent they achieve learning purposes.

In an assessment there is reliability, referring to assessment consistency which is done by teacher/lecturer. Kusaeri (2012:62-63) states consistency of a measuring activity is how the test score is consistent with measurement one to another. Based on the definition of reliability, some characteristics of reliability are: 1) reliability refers to the results which are obtained through a test instrument, not refers to the instrument itself; 2) reliability is a necessary requirement, but it is not sufficient to fulfill the requirement of validity. A test that is inconsistent in the result will not produce valid information related to the measured ability, so that low reliability can limit the level of validity which is obtained. Reliability produces a consistent result so that the validity will be fulfilled; 3) Reliability is related to statistic, logic analysis. To find out *inter – rater* consistency, the test must be conducted first. Consistency is usually stated in the form of reliability coefficient and *standard error of measurement*.

Assessment is an activity of intrepreting the result of measurement, for example succeced or fail, either good or bad, competent or incompetent, pass and fail, and the kind based on certain standard. Performance test in vocational education usually uses two approaches, i.e. holistic method and analytical method. Holistic method is used when the rater only gives one score (*single rating*), based on the whole assessment from the results of students' performance, while analytical method is used if the rater gives scores on various aspects related to the assessed performance. Analytical rubric is more detailed and it contains statements indicating the measured part or aspect. (Johnson, 2009: 119).

Performance assessment, including clothing making assessment uses assessment criteria, aiming at resulting understanding. The assessment referring to criteria is usually known as reference assessment. According to Djemari (2004:13), the main characteristics that mark the use of reference assessment is score intrepretation from the measurement instrument that can make a description regarding ability or knowledge which is owned by learners. The intrepretation of the test result is always compared to the standards or criteria which are determined previously.

The Final Project Course is a course that is obliged to be taken and passed for the students of Clothing Design as the final assessment in D3 Program, and as a subject for Undergraduate Program. There are some characteristics of the Final Project course, i.e. 1) as a plus practice course, because after a fashion product finished, it is shown in a fashion show, 2) the making process is started by creating design. In this course, the students are demanded to be more creative either in making design, making pattern, changing pattern, selecting materials, exploring materials, sewing technology, until appropriate to be shown in a fashion show.

The purpose of this research are: 1) to analyze score to get *Inter-Rater* consistency in the Final Project assessment year 2015 Clothing Design Education Study Program based on external assessment, 2) to analyze score to get consistency of *Inter-Rater* in the Final Project assessment year 2015 in the Clothing Design Education Study Program based on internal assessment.

## II. RESEARCH METHOD

This research is a survey research, i.e. to obtain an accurate decision from the rater or external and internal experts in the fashion show of the Final Project of Clothing Design Education Study Program year 2015. There are five raters or external experts who are from clothing practitioners, clothing academician, and clothing profession association. Meanwhile, the internal experts are advisors (lecturers of PT Busana) for each aspect consists of three raters. Population of this research is all students of the Clothing Design Education, while the research sample is the students at the sixth semester in Undergraduate Program. The sample technique applied is *Purposive Sampling,* i.e. the sixth semester students who are taking the Final Project course.

The data are collected by using assessment sheets which are obtained through score documents from the *rater* for all students. After that, the collected score is processed into the final score by using the determined formula that is based on the weight of each assessment aspect. The validity of the instrument that is arranged to present the research data needs to be known. It is obtained by asking opinions of the lecturers according to the field in which they work, i.e. design expert and clothing production expert, trimming expert through study program meeting forum. Then, before the external r*aters* (consists of 5 people) assess, there is a briefing to agree the aspects and criteria which will be assessed.

The data analysis technique applied in this research is descriptive analysis technique, i.e. explaining consistency of the final project measurement from external experts by using *Alpha* and ICC (*Intraclass Correlation*). The ICC analysis is aimed to assess the measuring consistency that is given by some observers who measure the same number.

## III. RESULTS AND DISCUSSION

After going through a long process for five months, the students finished the fashion products which were going to be presented in the fashion show. However, before the fashion products are shown, they must be assessed. The assessment of the fashion products were conducted to elect the winner by the external experts who were derived from practicians, clothing professional association, journalism, fashion analyst and education. The selection of board of expert is considered based on skill and experiences in fashion area.

Futhermore, the assessment activity is started by briefing of all external experts, aiming at consolidating perceptions in each aspect that will be measured. It is done to minimize level of subjectivity and to avoid any mistake during the assessment. Therefore, before the assessment activity is done, the form of assessment for each aspect must be understood and examined seriously to get understanding inter-raters. If there are some experts giving not really high score range, it is predicted that inter-raters have consistent or almost the same score.

### *A. External Assessment Consistency*

The assessment process is started by presenting a finished fashion product. The Final Project assessment's consistency is analyzed by using *Intraclass Correlatiom (ICC) and Alpha* techniques. The criteria of ICC coefficient is 0.50 at minimum, while Alpha coefficient at least 0.80. The ICC coefficient number is influenced by the range of the observed value (variability – variance), while the number of Alpha coefficient is influenced by the sample number. The result of Final Project assessment by the external raters is presented as follows.

Table 1. Summary of ICC and Alpha index from External Raters

| No | Study Program | The Assessed Aspects | | | | | | Inter-Rater Consistency Index | |
|---|---|---|---|---|---|---|---|---|---|
| | | Originality | | Look | | Design Appropriateness | | | |
| | | ICC | Alpha | ICC | Alpha | ICC | Alpha | ICC | Alpha |
| 1. | PT. Busana (A) | 0.31 | 0.69 | 0.26 | 0.63 | 0.30 | 0.68 | 0.29 | 0.67 |
| 2. | PT Busana (D) | 0.10 | 0.34 | 0.18 | 0.53 | 0.20 | 0.55 | 0.16 | 0.47 |

Based on the results of inter-rater consistency analysis from the external raters, it is shown that the coefficient of consistency based on ICC analysis < 0.50. The low ICC coefficient, there are some aspects which are assessed, i.e. originality, look, and design appropriateness. It is also shown for the coefficient of consistency based on Alpha analysis < 0.70. Thus, the consistency of the assessment for the three aspects show low consistency.

*B. Internal Assessment Consistency*

Internally, the assessment is conducted by the advisors. The assessed aspects are: trimming, clothing technology, design appropriateness, and design. Below is presented inter-rater consistency index, both using ICC or Alpha.

Table 2. Summary ICC and *Alpha* index from Internal Raters

| No | Study Program | The Assessed Aspects | | | | | | | | Average of *Inter-rater Consistency Index* | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Trimming | | Clothing Design | | Wearing Appropriateness | | Design | | | |
| | | ICC | Alpha | ICC | Alpha | ICC | Alpha | ICC | Alpha | ICC | Alpha |
| 1. | PT. Busana (A) | 0.26 | 0.51 | 0.62 | 0.83 | 0.57 | 0.82 | 0.67 | 0.86 | 0.53 | 0.76 |
| 2. | PT Busana (D) | 0.45 | 0.71 | 0.62 | 0.83 | 0.25 | 0.50 | 0.42 | 0.68 | 0.44 | 0.68 |

The results of *inter-rater* analysis consistency from the internal raters (advisors) show that the coefficient of consistency based on ICC analysis < 0.50 is only found in trimming assessment consistency. The aspects of clothing technology, wearing appropriateness and design have fulfilled the required ICC coefficient that is > 0.50. The Alpha coefficient shows that trimming aspect and clothing technology aspect show good consistency. Meanwhile, for other aspects which are wearing appropriateness and design show coefficient < 0.7. It is concluded that the inter-rater consistency is low.

Based on the results of consistency based on *Intraclass Correlatiom (ICC) and Alpha*, it shows that external inter-raters do not show high level of understanding or consistency or similarity in scoring the results of the students' fashion products. After doing deep observation, it is found that there is incomplete instrument, such as rubric. Although there are other instruments, without clear rubric it will make the assessment in determining score unclear. If the assessment instrument is available before the assessment is conducted, it will make the raters easier in determining scores. The number of the inter-rater scores which are almost the same will decide the actual students' competency. Thus, in determining the winner, there are scores which are almost the same or consistent. If the students get similar scores or almost the same, it will be accepted sincerely without disappointed, mad, sad and unfair feelings and so on toward the final products.

## IV. CONCLUSION AND SUGGESTION

It can be concluded that:1. Based on the results of inter- rater consistency analysis from external raters, it is shown that the coefficient of consistency is based on ICC analysis < 0.50, while the coefficient of consistency based on Alpha analysis < 0.80. Based on the results of inter rater consistency analysis from the internal raters (advisors), it shows that the coefficient of consistency based on ICC analysis < 0.50, while the coefficient of consistency based on Alpha analysis < 0.80. However, the interrater consistency of each aspect shows good consistency.

Some important suggestions can be concluded are: 1) regarding the low consistency coefficient of the external experts, either by using ICC coefficient or Alpha coefficient, for the next fashion show, it needs to be selective in determining experts, to get decision regarding the students' works accurately; 2) in order to get good consistency of the next Final Project, complete and clear assessment instruments are needed.

## REFERENCES

[1] M Djemari, **"**Pengembangan sistem penilaian berbasis kompetensi," in *Proceeding*: Rekayasa sistem penilaian dalam rangka meningkatkan kualitas pendidikan, Yogyakarta: HEPI, 2004.

[2] Kusaeri, Suprananto, "Pengukuran dan penilaian pendidikan," Yogyakarta: Graha Ilmu, 2012.

[3] R.L. Johnson, J.A. Penny, B. Gordon, "Assessing performance: designing, scoring, and validating performance task," London: The Guilford Pres, 2009.

[4] S. A. Yorkovich, G.S. Waddell, R.K. Gerwig, "Competency-based assessment systems: Encouragement toward a more holistic approach," *http://spiritoforganization.com/documents/Waddell_CompetencyBasedAssessment.pdf,* 2008. (Date of access: 5 Januari 2010)