

ISSN: 2407-1501

PROCEEDING

INTERNATIONAL CONFERENCE ON EDUCATIONAL RESEARCH AND EVALUATION (ICERE)

“Authentic Assessment for improving Teaching Quality”

November 8-9, 2014
Rectorate Hall and Graduate School
Yogyakarta State University
Indonesia





**Organized by:
Doctoral and Master's Program in
Educational Research and Evaluation
Graduate School, Yogyakarta State University**

Proceeding

International Conference on Educational Research and Evaluation (ICERE) 2014

Publishing Institute

Yogyakarta State University

Director of Publication

Prof. Djemari Mardapi, Ph.D.

Board of Reviewers

Prof. Djemari Mardapi, Ph.D.

Prof. Dr. Badrun Kartowagiran

Prof. Dr. Sudji Munadi

Prof. Dr. Trie Hartiti Retnowati

Dr. Heri Retnowati

Dr. Widiastuti

Editors

Ashadi, Ed.D.

Suhaini M. Saleh, M.A.

Titik Sudartinah, M.A.

Lay Out

Anggit Prabowo, M.Pd.

Rohmat Purwoko, A.Md.

Address

Yogyakarta State University

ISSN: 2407-1501

@ 2014 Yogyakarta State University

All right reserved. No part of this publication may be reproduced without the prior written permission of Yogyakarta State University

All articles in the proceeding of International Conference on Educational Research and Evaluation (ICERE) 2014 are not the official opinions and standings of editors. Contents and consequences resulted from the articles are sole responsibilities of individual writers.

BACKGROUND

In its effort to improve the quality of education in Indonesia, the Indonesian government has imposed Curriculum 2013 on schools of all level in Indonesia. The main difference between Curriculum 2013 and the previous curriculum lies in its implementation which uses the scientific approach. For the reason, teachers need to develop teaching strategies different from those they used to apply in the implementation of the previous curriculum. Besides, teachers also need to develop the techniques of evaluating students' learning achievement, which are relevant to the scientific approach. The evaluation has to be able to show the students' learning achievement in observing, experimenting, social networking, etc.

Authentic assessment conducted in the classroom and focusing on complex and contextual tasks enables students to perform their competence in a more authentic arrangement. It is very relevant to the authentic approach integrated in their teaching process, especially at elementary schools, or for appropriate lessons. It must be able to show which attitude, skill, and knowledge have or have not been mastered by the students, how they use their knowledge, what aspect they have or have not been able to apply, and so on.

On the basis of the above consideration, teachers can identify what materials the students can study further and for what material they need to have a remedial program. Authentic assessment, however, is not that easy!

FOREWORD

In the academic year of 2014, the government in this case the Ministry of Education and Culture has established the policy to run the curriculum of 2013 for the all levels of elementary and intermediate education in Indonesia. It means the schools have to be ready to implement the Curriculum of 2013. Basically, the implementation of the 2013 curriculum is an effort from the government to enhance the quality of education.

One of the characteristics of the 2013 curriculum is make use the scientific approach in the learning process. This approach is to improve the students' creativity in learning. In general, this approach seems to be a new thing for the teachers in which several problems and obstacles appear in its practice. The teachers are required to develop the learning strategies and the assessment systems which are relevant and appropriate in order to nurture the students' creativity. One of the assessment methods that can support the concept of scientific approach is by using the authentic assessment. Authentic assessment can give the description of the knowledge, the attitudes, and the skills as well as what has or has not owned by the students and the way they apply their knowledge. Also, in what case they have or have not been able to implement the learning acquisition.

Based to the above circumstances, the Study Program of Educational Research and Evaluation, Graduate School of Yogyakarta State University (Universitas Negeri Yogyakarta) conduct the international seminar on the theme "Classroom Assessment for Improving Teaching Quality". There will be three sub-themes on this seminar, i.e. Issues of Classroom Assessment Implementation, Implementation of Authentic Assessment, and Developing a Strategy of Creative Teaching. By having this seminar, the participants are expected to possess the knowledge and the skills to develop and to apply the authentic assessment.

Yogyakarta, November 8, 2014
Head of Committee

Prof. Dr. Sudji Munadi

CONTENTS:

	pages
Title	i
Background	ii
Foreword	iii
Welcome Speech	iv
Preface	v
Content	vi

Invited Speaker

ISSUES OF CLASSROOM ASSESSMENT IMPLEMENTATION Madhabi Chatterji	2
IMPLEMENTATION OF AUTHENTIC ASSESSMENT Pongthep Jiraro.....	11
DEVELOPING A STRATEGY OF CREATIVE TEACHING Paulina Panen.....	23

Paper Presenter

Theme 1:

Issues of Classroom Assessment Implementation

ASSESSMENT IN DEVELOPMENT COMPUTER-AIDED INSTRUCTION Abdul Muis Mappalotteng	35
THE MEASUREMENT MODEL OF INTRAPERSONAL AND INTERPERSONAL SKILLS CONSTRUCTS BASED ON CHARACTER EDUCATION IN ELEMENTARY SCHOOLS Akif Khilmiyah.....	47
LEARNING ASSESSMENT ON VOCATIONAL SUBJECT MATTERS OF THE BUILDING CONSTRUCTION PROGRAM OF THE VOCATIONAL HIGH SCHOOL IN APPROPRIATE TO CURRICULUM 2013 Amat Jaedun	63
ACCURACY OF EQUATING METHODS FOR MONITORING THE PROGRESS STUDENTS ABILITY Anak Agung PurwaAntara	74

EFFECT OF PERFORMANCE ASSESSMENT ON STUDENTS' THE ACHIEVEMENT IN PHYSICS HIGH SCHOOL Aswin Hermanus Mondolang	86
TEST ITEM ANALYSIS PROGRAM DEVELOPMENT WITH RASCH MODEL ONE PARAMETER FOR TESTING THE ITEM DIFFICULTY LEVEL OF MULTIPLE-CHOICE TEST USING BLOODSHED DEV C++ APPLICATIONS Dadan Rosana, Otok Ewi Amsirta	93
EFFECTIVENESS OF REASONED OBJECTIVE CHOICE TEST TO MEASURE HIGHER ORDER THINKING SKILLS IN PHYSICS IMPLEMENTING OF CURRICULUM 2013 Edi Istiyono, Djemari Mardapi, Suparno	101
DEVELOPING STUDENTS' SELF-ASSESSMENT AND STUDENTS' PEER-ASSESSMENT OF THE SUBJECT-MATTER COMPETENCY OF PHYSICS EDUCATION STUDENTS Enny Wijayanti, Kumaidi, Mundilarto	111
THE RESULT OF ASSESSMENT FOR STUDENTS IN SOLVING EXPONENTS AND LOGARITHMS PROBLEMS (CASE STUDY IN GRADE X CLASS MATHEMATICS AND NATURAL SCIENCE (MIA) 2 STATE SENIOR HIGH SCHOOL 1 DEPOK 2014/2015) Fajar Elmy Nuriyah	123
RELIABILITY RANKING AND RATING SCALES OF MYER AND BRIGGS TYPE INDICATOR (MBTI) Farida Agus Setiawati	131
THE COMPARISON OF ITEMS' AND TESTEES' ABILITY PARAMETER ESTIMATION IN DICHOTOMOUS AND POLITOMUS SCORING (STUDIES IN THE READING ABILITY OF TEST OF ENGLISH PROFICIENCY) Heri Retnawati	139
STUDENTS' CHARACTER ASSESSMENT AS A REFERENCE IN TEACHING LEARNING PROCESS AT SMPK GENERASI UNGGUL KUPANG KorneliusUpa Rodo, Netry E.M. Maruckh, Joko Susilo	151
MEASUREMENT ERROR ESTIMATION OF CUT SCORE OF ANGOFF METHOD BY BOOTSRA TP METHOD Sebastianus Widanarto Prijowuntat	157

THE ACTUALIZATION OF PROJECT-BASED ASSESSMENT IN ENTREPRENEURSHIP EDUCATION BASED ON LOCAL EXCELLENCE IN MEASURING SKILLS OF VOCATIONAL HIGH SCHOOL STUDENTS Sukardi	168
THE EFFECTIVENESS OF THE USE OF THE INSTRUMENTS AND RUBRICS OF CREATIVE THINKING SKILLS–BASED ASSESMENT PROJECT IN THE LEARNING OF CONSUMER EDUCATION Sri Wening	177
PROJECT WORK USED IN A COMPREHENSIVE ASSESSMENT TO MEASURE COMPETENCES OF UNDERGRADUATE ENGINEERING STUDENTS Sudiyatno.....	194
THE DEVELOPMENT OF A SET OF INSTRUMENT FOR STUDENT PERFORMANCE ASSESSMENT Supahar	202
DEVELOP MODEL TASC TO IMPROVE HIGHER ORDER THINKING SKILLS IN CREATIVE TEACHING Surya Haryandi	210
THE EFFECT OF NUMBER’S ALTERNATIVE ANSWERSON PARTIAL CREDIT MODEL (PCM) TOWARDESTIMATION RESULT PARAMETERS OF POLITOMUS ITEM TEST Syukrul Hamdi	216
THE CONTENT VALIDITY OF THE TEACHER APTITUDE INSTRUMENT Wasidi	227
DEVELOPING COGNITIVE DIAGNOSTIC TESTS ON LEARNING OF SCIENCE Yuli Prihatni	233
DIAGNOSTIC MODEL OF STUDENT LEARNING DIFFICULTIES BASED ON NATIONAL EXAM Zamsir, Hasnawati	246
DEVELOPMENT OF A MODEL OF ACADEMIC ATTITUDE AMONG SENIOR HIGH SCHOOL STUDENTS Sumadi	258

Theme 2:

Implementation of Authentic Assessment

IMPLEMENTATION OF AUTHENTIC ASSESSMENT OF CURRICULUM 2013 AT STATE ELEMENTARY SCHOOLS IN PABELAN Abdul Mu'in, Nining Marianingsih, Woro Widyastuti	265
AUTHENTIC ASSESSMENT OF STUDENT LEARNING MATHEMATICS WITH TECHNOLOGY Ida Karnasih	278
AUTHENTIC ASSESSMENT : UNDERSTANDING LEVELS AND CONSTRAINTS IN THE IMPLEMENTATION OF THE TEACHER IN THE CITY OF LHOKSEUMAWE ACEH PROVINCE M. Hasan	286
AUTHENTIC ASSESSMENT DETERMINANT IN ISLAMIC RELIGION EDUCATION EXECUTION TOWARDS COGNIZANCE QUALITY HAVES A RELIGION IN STUDENT AT ELEMENTARY SCHOOL AND MADRASAH IBTIDAIYAH AT KUDUS REGENCY Masrukhin	295
AUTHENTIC ASSESSMENT FOR IMPROVING TEACHING QUALITY: PORTFOLIO AND SLC IN PAPUA HARAPAN SCHOOL Noveliza Tepy, Sabeth Nuryana, Putri Adri	307

Theme 3:

Developing a Strategy of Creative Teaching

THE EFFECT OF MATH LESSON STUDY IN TERMS OF MATHEMATICS TEACHER'S COMPETENCE AND MATH STUDENT ACHIEVEMENT 'Afifatul Muslikhah	319
AN EVALUATION OF THE ENGLISH TEACHING METHODS IMPLEMENTED AT BUJUMBURA MONTESSORI PRIMARY SCHOOL: WEAKNESSES AND ACHIEVEMENTS Alfred Irambona	323
TEAMS GAME TOURNAMENT FOR IMPROVING THE STUDENTS' INTEREST TOWARD MATHEMATICS Anggit Prabowo	331

DEVELOPING LEARNING KIT TO IMPROVE HOTS FOR FLAT SIDE OF SPACE COMPETENCE Arifin Riadi	346
DEVELOPMENT STRATEGY OF TEACHERS' TEACHING PROFESSIONALISM Bambang Budi Wiyono	352
THE EFFECT OF QUESTION PROMPTS AND LANGUAGE ABILITY ON THE QUALITY OF THE STUDENT'S ARGUMENT Bambang Suteng Sulasmono, Henny Dewi Koeswanti	364
THE USE OF RESPONSE ACTIVITIES IN DEVELOPING READING SKILLS AMONG INTERMEDIATE EFL STUDENTS Beatriz Eugenia Orantes Pérez	378
COMPARISON OF THE EFFECTIVENESS OF CONSTRUCTIVISM AND CONVENTIONAL LEARNING KIT OF MATHEMATICS VIEWED FROM ACHIEVEMENT AND SELF CONFIDENCE OF STUDENTS IN VOCATIONAL HIGH SCHOOL (AN EXPERIMENTAL STUDY IN YEAR XI OF SMK MUHAMMADIYAH 2 YOGYAKARTA) Dwi Astuti, Heri Retnawati	389
THE EFFECT OF CLASS-VISITATION SUPERVISION OF THE SCHOOL PRINCIPAL TOWARD THE COMPETENCE AND PERFORMANCE OF PANGUDI LUHUR AMBARAWA ELEMENTARY SCHOOL TEACHERS Dwi Setiyanti, Lowisye Leatomu, Ari Sri Puranto, Theodora Hadiastuti, Elsavior Silas	394
THE 'REOP' ARCHITECTURE TO IMPROVE STUDENTS LEARNING CAPACITY Edna Maria, Febriyant Jalu Prakosa, Christiana, Monica Ganeip Pertiwi	403
E-LEARNING-BASED TRAINING MODEL FOR ACCOUNTING TEACHERS IN EAST JAVA Endang Sri Andayani, Sawitri Dwi Prastiti, Ika Putri Larasati, Ari Sapto ...	408
CONCEPT AND CONTEXT RELATIONSHIP MASTERY LEARNING AND THE RELATIONSHIP BETWEEN BIOLOGY AND PHYSICS CONCEPT ABOUT MANGROVE FOREST Eva Sherly Nonke Kaunang	431

THE EFFECTIVENESS OF TEACHING MULTIMEDIA ON TOPIC OF THREE DIMENSIONS IN TERMS OF THE MATHEMATICS LEARNING ACHIEVEMENT AND INTEREST OF STATE SENIOR HIGH SCHOOL Lisner Tiurma, Heri Retnawati	439
BUILDING THE STUDENT CHARACTER THROUGH THE ACADEMIC SERVICE M. Miftah	448
THE TEACHING EVALUATION OF GERMAN TEACHER IN MALANG Primardiana Hermilia Wijayati	464
SUPPORTING PHYSICS STUDENT LEARNING WITH WEB-BASED ASSESSMENT FOR LEARNING Sentot Kusairi, Sujito	479
AMONG LEARNING AS A CULTURE BASED LEARNING OF TAMAN MUDA TAMAN SISWA AS CONTRIBUTION TO THE LEARNING PROCESS OF 2013 CURRICULUM AND CHARACTER EDUCATION OF THE NATION Siti Malikhah Towaf	490
THE PERFORMANCE OF THE BACHELOR EDUCATION IN-SERVICE TEACHERS PROGRAMME (ICT-BASED BEITP) BACHELOR GRADUATED AND ITS DETERMINANT Slameto	513
DEVELOPING LEARNING TOOLSOFA GAME-BASED LEARNING THROUGH REALISTIC MATHEMATICS EDUCATION (RME) FOR TEACHING AND LEARNING BASED ON CURRICULUM 2013 Sunandar, Muhtarom, Sugiyanti	525
PREPARATION OF COMPUTER ANIMATION MODEL FOR LEARNING ELECTRICAL MAGNETIC II PHYSICAL EDUCATION PROGRAM STUDENTS SEMESTER IV TEACHER TRAINING AND EDUCATION FACULTY SARJANAWIYATA TAMANSISWA UNIVERSITY 2014 Sunarto	536
IMPROVEMENT ACTIVITIES AND STUDENT LEARNING OUTCOMES IN READING COMPREHENSION THROUGH COOPERATIVE LEARNING TYPE TEAMS-GAMES-TOURNAMENT (TGT) CLASS V SD NEGERI 8 METRO SOUTH Teguh Prasetyo, Suwarjo, Sulistiasih	547

PSYCHOLOGICAL FACTOR AFFECTING ENGLISH SPEAKING PERFORMANCE FOR THE ENGLISH LEARNERS IN INDONESIA Youssof Haidara	558
---	-----

RELIABILITY RANKING AND RATING SCALES OF MYER AND BRIGGS TYPE INDICATOR (MBTI)

Farida Agus Setiawati
faridaagus@yahoo.co.id.

Abstract

One of the personality models is Myers Briggs Type Indicator (MBTI). This model has been developed in the personality instrument using force choice and has been administration by the simple format. The base of force choice is ranking format. It has weakness completed by rating format. The study aims to analyze the reliability of MBTI personality using ranking and rating format, and to present the characteristic of personality undergraduate student using two format.

The study employed quantitative methods. The subject of the research is 134 undergraduate students of Yogyakarta State University. They come from three study programs, the educational of math, the educational of economic, and guidance and counseling. The data were collected by giving questionnaire. The reliability of each dimensions instrument was analyzed using alpha Cronbach. The all dimensions were analyzed using composite reliability. The personality characteristics subject was analyzed by descriptive statistic.

The findings of this study show that two instrument were reliable. The reliability instruments were various in many dimensions. The lowest reliability was sensing type and the highest reliability was thinking and feeling in the ranking format, and the extrovert in the rating format. The lowest reliability was the sensing not only in the ranking but in the rating format as well. The all reliability instrument or composite reliability of ranking format was 0.74 and the rating format was 0.72, It showed that the reliability of forced-choice format was higher than rating format. The result of measurement personality characteristic of Yogyakarta undergraduate student is dominant extraversion in focus attention, dominant sensing in finding out about things, dominant feeling in making decision and dominant judging in orientation the other world. Consequently, Yogyakarta undergraduate students prefer to outer world, strive for harmony with others and friendship, and prefer to do something ordered and scheduled.

Keyword: *personality, reliability, ranking, rating, MBTI*

Introduction

The life will be happy if we know our personalities. Personality is a pattern of the way of thinking, feeling, showing character, and conducting activity that determining person's ability to adapt to the environment. One of the instruments identifying characteristic personality of person is Myers-Briggs Type Indicator or MBTI (Boyd & Brown, 2005; Bradley & Hebert, 1997). The instrument was developed by Katharine Cook Briggs and her daughter, Isabel Briggs Myers, based on Carl Gustav Jung's theory of Psychological Types

(1921/1971). Jung argues that there are four basic psychological functions of human related to the world, namely: sensation, intuition, feel, and thought (Briggs-Myers & McCaulley, 1985). They argue that the characteristics of human behavior can be distinguished on the types of preferences. These preferences are based on: 1) focusing direction consisting of extravert and introvert types; 2) the way of obtaining information consisting of sensing and intuition type; 3) the way of making decisions consisting of thinking and feeling type; and 4) the orientation towards the outside world consisting preferences of judging and perceiving.

MBTI instrument is formed by paired comparison format. This format was designed from pairing statement in one item. From the two alternative statements in each item, it must be selected only one statement from the two types of the same dimension. For example, the introvert type is paired with the extrovert type, the thinking type is paired with the feeling type, the sensing type is paired with the intuition type, and the judging type is paired with the perceiving type. The paired comparison format is introduced by Thurstone (1927). This format is based on the forced-choice format (McDonald, 1999). It was the measurement model that forces the subject to response one of two or more statements.

The forced-choice is same as ranking format if there are two responses or statement that must be chosen. This format has two advantages to the response of a subject. It is avoided from social desirability and faking (McDonald, 1999, Barislow, 1958, Chernyshenko, 2009). Social desirability is a general statement in which people possibly respond to the statement dishonestly. It may be concluded that subject's response is known as faking. Therefore, it is difficult to detect the real opinion or response from the subject. Nevertheless, this type of instrument has weakness because the subject is forced to choose one statement though both of them are appropriate or inappropriate. On the other side, another model instrument, called the rating model, is responded freely. Subjects are free to respond to the statement as they wish though they may respond to it by faking. Thus, the weakness of forced-choice or ranking format can be covered by the rating format, and vice versa.

The comparison of the two formats of instrument has been studied by Setiawati (2013). She studied the scaling of multiple intelligence instrument using Thurstone types or ranking format and Likert type designed by a rating format. The results show that Likert-type instrument is more accurate than the Thurstone type is. The accuracy is caused by Likert type's variations of the answer which is more than the Thurstone type is. The Likert type has five responses while the Thurstone type has two responses. Thus, it needs further research to determine the accuracy of both types by creating the same variations.

The result of measurement using MBTI instrument is characteristic personality from several pattern of personality from many types. There are four dimensions of the two type of personality. Thus, the all patterns of personality have 16 alternate personality types. MBTI is widely used to determine differences in personalities in various fields, such as education, career development, organizational behavior, group functions, team development, personal and executive training, individual psychotherapy, couples, and families, and context of multicultural interaction (Center for Applications of Psychological Type, 2008). It design originally by ranking format, but it can be design by rating format. The Data of this study were collected using two format instrument.

This study aims to investigate the characteristics of the instrument MBTI and the types of personality of students in Yogyakarta by using the format type instrument with ranking and rating.

Method

The data were collected by giving questionnaire of MBTI instrument. It was modified into two formats, the forced-choice format (the original format) and the rating format. Both of format, were formed from the same construct and item. The specification items of the instrument can be showed in the table 1.

Table 1 Specification Items of the Instrument

Based on Preferences	Types of Preference	Number of Items	
		Forced-choice format	Rating format
Focusing direction	Extrovert (E)	1,9,19,29,35,41,53	1A,5A,10A,15A,18A,21A,27A
	Introvert (I)	2,10,20,30,36,42,54	1B,5B,10B,15B,18B,21B,27B
The way of obtaining information	Sensing (S)	3,13,17,31,39,45,55	2A,7A,9A,16A,20A,23A,28A
	Intuition (N)	4,14,18,32,40,46,56	2B,7B,9B,16B,20B,23B,28B
The way of making decisions	Thinking (T)	7,11,21,25,37,47,49	4A,6A,10A,13A,19A,24A,25A

	Feeling (F)	8,12,22,26,38,48,50	4B,6B,10B,13B,19B,24B,25B
The orientation towards the outside world	Judging (J)	5,15,23,27,33,43,51	3A,8A,12A,14A,17A,22A,26A
	Perceiving (P)	6,16,24,28,34,44,52	3B,8B,12B,14B,17B,22B,26B

From table 1, could be showed that both of two format have the same items but different in the number of item. Then, they were tested to subject of research. The subject of the research is undergraduate students of Yogyakarta State University. Meanwhile, to determine subject of research, the cluster random sampling was employing. They come from three study programs, the educational of math, the educational of economic, and guidance and counseling. The all subject were 134 undergraduate students.

Finally, there were two techniques to analyze the data: 1) analyzed the reliability instrument, 2) analyzed the characteristic of personality. The reliability of each dimensions instrument was analyzed using alpha Cronbach. Then, the all dimensions were analyzed by using composite reliability. The characteristic of personality subject was quantitative analyzed by descriptive statistic. All analysis of the data used Excel program.

Results

One of the characteristic instruments in classical theory was reliability. The concept of reliability in classical test theory assumptions is associated with observed scores (X), true score (T), and error score (E). The main concept of this theory is $X = T + E$, or the observed score is the combination of the true scores and error scores. Based on these assumptions, the concept of variant scores or total variance is the combination of the true score variance and the error score variance. The assumptions made in the formula $\sigma_t^2 = \sigma_r^2 + \sigma_e^2$ (Gulliksen, 1950, Lord & Novick, 1968, Allen & Yen, 1979, Thissen & Wainer, 2001).

The reliability of data instrument associated with measurement errors in the measurement data. Djemari Mardapi (2008) said that measurement error is part of unreliability data. It is studied in measurement of social sciences. Based on the concept of scores in the classical theory, the reliability of the measured data can be explained from the variant score. An association between the variant score as mentioned in the assumptions of classical theory can be used to explain the definition of reliability which is a variant of

interaction errors and score. The concept of reliability can be formulated as $\rho_{xx'} = 1 - \sigma_e^2 / \sigma_x^2$, where reliability is the magnitude of the error variance and variance score. Based on this formula, it can be explained that the larger the error variance is the smaller the coefficient reliability and will be vice versa. The score of reliability, variance score, variance error influence to the characteristic of instrument. The formula alpha from Cronbach was used to estimate the reliability from two formats. The result of data analyze could be seen in table 2.

Table 2. The Reliability and Standard Error Measurement (SEM) of Two Formats Instrument

No	Dimension	Ranking-format			Rating-format		
		Variance	Reliability	SEM	Variance	Reliability	SEM
1.	Extrovert	2,66	0,53	1,83	2,72	0,6	1,1
2.	Introvert	2,66	0,53	1,83	2,45	0,43	1,39
3.	Sensing	1,85	0,3	1,55	1,88	0,34	1,24
4.	Intuition	1,85	0,3	1,55	2,78	0,52	1,34
5.	Judging	3,09	0,6	1,94	2,71	0,5	1,34
6.	Perceiving	3,09	0,6	1,94	2,35	0,42	1,36
7.	Thinking	3,52	0,62	2,17	2,88	0,57	1,24
8.	Feeling	3,52	0,62	2,17	2,62	0,57	1,14
9.	Total instrument		0.74	1,87	31	0.72	1,27

Table 2 shows that the reliability instruments were various in many dimensions. The lowest reliability was sensing type and the highest reliability was thinking and feeling in force-choice type, and the extrovert type in the rating format. The lowest reliability was the sensing not only in the force-choice format but in the rating format as well. The all reliability instrument or composite reliability of ranking format was 0.74 and the rating format was 0.72. It showed that the reliability of forced-choice format was higher than rating format.

The data from forced-choice and rating format are different. Although the data from both types of instruments are different but there are connected. The calculation of coefficient correlation from many dimension or two formats are calculated using Pearson correlation.

The result of correlation test shows that coefficient correlation or $r = 0.791$ and probability of significance or $p = 0.02$. Consequently, the two of formats are connected.

The term of error measurement related to the error variance or standard error of measurement (SEM). The formula of SEM or $\sigma_e = \sigma_t \sqrt{1 - \rho_{xx'}}$. SEM magnitude affects the reliability, it can be said that the accuracy of measurement results can be seen by the SEM. The smaller the SEM measurement will be more precise measurement results. Table 3 showed that the SEM of ranking format was higher than rating format, therefore, unless the ranking format have more reliable than the rating format, the rating format had more precise. It condition was influenced of the variance of ranking format was higher than the rating format.

The data from ranking and rating format are different. Although the data from both types of instruments are different but there are connected. The calculation of coefficient correlation from many dimension or two formats are calculated using Pearson correlation. The result of correlation test show that index correlation or $r = 0.791$ and probability of significance or $p = 0.02$. Consequently, the two of formats are connected.

The Profile of characteristic personality was analyzed by descriptive statistic using mode. The mode data from two format instrument could be seen in table 2.

Table 2 Descriptive Statistic the Data of Two Format Instrument

No	Dimension	Mode	
		Force-choice	Rating Formal
1	Extrovert	4	5
2	Introvert	3	5
3	Sensing	5	6
4	Intuition	2	4
5	Judging	5	6
6	Perceiving	2	5
7	Thinking	3	4
8	Feeling	4	5

Table 2 shows that the preferences of focusing direction, in force-choice format the extrovert is more dominant than introvert is, but in the rating format both types have similar score. Another score shows that to obtaining information, in the two formats instrument, the sensing is more dominant than intuition is. Meanwhile, the way of making decisions, the feeling is more dominant than the thinking is in both of formats instrument, and the orientation towards the outside world, the judging is more dominant than perceiving is in the two formats. From the all preferences, the characteristic personality of subjects are extrovert, sensing, judging and feeling (ESJF). Consequently, the pattern of personality of Yogyakarta State University Students is ESJF.

The extrovert people have motivated and strong energy from outside and environment. The sensing people like taking information from their perception or sensation. The judging people have many plans and arrangements in their live. And the feeling people have many value in their lives and base on humanity in the interaction with the other people. From this characteristic, Yogyakarta undergraduate students prefer to outer world, prefer to do something ordered and scheduled, and strive for harmony with others and friendship.

The score of rating format is higher than ranking format is. It is coursed that the rating format might be to response high in the pairs of items. Choosing one statement out of several statements in ranking makes this type of instrument have difference characteristic to other instruments which provide a statement with various different responses. This instrument will be different responses for one stimulus. Hence, it will be obtained various kinds of responses on several provided stimuli. This is inline with Oliveres & Brown's (2010, p.935) stated that this type of instrument is spared from the same answers or there is bias in giving responses such as an extreme agreement response. It is also stated that by using this type of instrument, weaknesses in responding such as lack of various responses or 'halo effect' will be avoided.

The ranking instrument has strength related to subject response that tends to be spared from social desirability and faking (Mc Donald, 1999, p.24, Chernyshenko, et.al., 2009, p.108). Therefore, Social desirability is a general statement that tends to be responded by a subject dishonestly, that creates difficulties in finding out the real opinion of the subject. Meanwhile, *faking* is a tendency of subjects to choose responses that do not represent their characteristics or intentionally change their responses to make them included in the desired group. Thus, both formats instrument have weakness and strength in their object matter.

Conclusion

The findings of this study show that two formats of instrument are reliable in the both formats instrument using composite reliability. The reliability instruments were various in many dimensions. The lowest reliability was sensing type and the highest reliability was thinking and feeling in the ranking format, and the extrovert in the rating format. The lowest reliability was the sensing not only in the ranking but in the rating format as well. The all reliability instrument or composite reliability of ranking format was 0.74 and the rating format was 0.72, It showed that the reliability of ranking format was higher than rating format.

The result of measurement personality characteristic of Yogyakarta undergraduate student is dominant extraversion in focus attention, dominant sensing in finding out about things, dominant feeling in making decision and dominant judging in orientation the other world. Consequently, Yogyakarta undergraduate students prefer to outer world, strive for harmony with others and friendship, and prefer to do something ordered and scheduled.

References

- Allen, M. J., & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey : Brooks/Cole Publishing Company.
- Borislw, B. (1958). The Edwards personal preference schedule (EPPS) and fakability. *Journal of Applied Psychology*, 42,1, 22-27
- Boyd, R., & Brown, T. (2005). *Pilot study of Myers Briggs Type Indicator personality profiling in emergency department senior medical staff*. Emergency Medicine Australia, 17
- Briggs-Myers, I.B. & McCaulley, M.H. (1985). *Manual: A Guide to the Development and Use of The Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.
- Chernyshenko, O. S., Stark,S., Prewett, M., Gray, A.A., Stilson, F. R.,& Tuttle,M. D. (2009). Normative scoring of multidimensional pairwise preference personality scales using IRT: empirical comparisons with other formats. *Human Performance*, 22, 105–127.
- Djemari Mardapi. (2008). *Teknik penyusunan instrumen tes dan nontes*.Yogyakarta: Mitra Cendekian Offset.
- Gulliksen, H., (1950). *Theory of mental tests*. New York: John Wiley and Sons. Inc.
- Jung, C. (1971). *Psychological types* (H. G. Baynes, Trans., revision by R. F. C. Hull), *The Collected Works of C. G. Jung* (Vol. 6), Princeton, NJ: Princeton University Press, (Original work published 1921).

- Lord, F. M., & Novick, M. R. (1974). *Statistical theories of mental test scores*. New York: Addison-Wesley Publishing Company, Inc.
- McDonald R. P. (1999). *Test theory: a unified threament*. London: Lawrence Erlbaum Associates.
- McLellan, C. R. (2011). Differences in Myers-Briggs personality types among high school band, orchestra, and choir members. *Journal of Research in Music Education*, 59(1), 85-100.
- Olivares, A. M., & Brown, A. (2010). Item response modeling of paired comparison dan ranking data. *Multivariate Behavioral Research*, 45, 935-974.
- Thissen, & Weiner, D. H. (2001). *Test scoring*. Marwah: Lawrence Erbaum Associated.
- Thurstone, L. L. (1927). Method of paired comparison for social values. *Journal of Abnormal and Social Psychology*, 21, 384-400.
- Setiawati, F. A. (2013). Perbandingan karakteristik psikometrik penskalaan tipe Likert dan Thurstone dengan teori klasik dan modern, studi pada instrumen MI. *Dissertation*. Yogyakarta State University.

RELIABILITY RANKING AND RATING SCALES OF MYER AND BRIGGS TYPE INDICATOR (MBTI)

Farida Agus Setiawati
faridaagus@yahoo.co.id.

Abstract

One of the personality models is Myers Briggs Type Indicator (MBTI). This model has been developed in the personality instrument using force choice and has been administration by the simple format. The base of force choice is ranking format. It has weakness completed by rating format. The study aims to analyze the reliability of MBTI personality using ranking and rating format, and to present the characteristic of personality undergraduate student using two format.

The study employed quantitative methods. The subject of the research is 134 undergraduate students of Yogyakarta State University. They come from three study programs, the educational of math, the educational of economic, and guidance and counseling. The data were collected by giving questionnaire. The reliability of each dimensions instrument was analyzed using alpha Cronbach. The all dimensions were analyzed using composite reliability. The personality characteristics subject was analyzed by descriptive statistic.

The findings of this study show that two instrument were reliable. The reliability instruments were various in many dimensions. The lowest reliability was sensing type and the highest reliability was thinking and feeling in the ranking format, and the extrovert in the rating format. The lowest reliability was the sensing not only in the ranking but in the rating format as well. The all reliability instrument or composite reliability of ranking format was 0.74 and the rating format was 0.72, It showed that the reliability of forced-choice format was higher than rating format. The result of measurement personality characteristic of Yogyakarta undergraduate student is dominant extraversion in focus attention, dominant sensing in finding out about things, dominant feeling in making decision and dominant judging in orientation the other world. Consequently, Yogyakarta undergraduate students prefer to outer world, strive for harmony with others and friendship, and prefer to do something ordered and scheduled.

Keyword: *personality, reliability, ranking, rating, MBTI*

Introduction

The life will be happy if we know our personalities. Personality is a pattern of the way of thinking, feeling, showing character, and conducting activity that determining person's ability to adapt to the environment. One of the instruments identifying characteristic personality of person is Myers-Briggs Type Indicator or MBTI (Boyd & Brown, 2005; Bradley & Hebert, 1997). The instrument was developed by Katharine Cook Briggs and her daughter, Isabel Briggs Myers, based on Carl Gustav Jung's theory of Psychological Types

(1921/1971). Jung argues that there are four basic psychological functions of human related to the world, namely: sensation, intuition, feel, and thought (Briggs-Myers & McCaulley, 1985). They argue that the characteristics of human behavior can be distinguished on the types of preferences. These preferences are based on: 1) focusing direction consisting of extravert and introvert types; 2) the way of obtaining information consisting of sensing and intuition type; 3) the way of making decisions consisting of thinking and feeling type; and 4) the orientation towards the outside world consisting preferences of judging and perceiving.

MBTI instrument is formed by paired comparison format. This format was designed from pairing statement in one item. From the two alternative statements in each item, it must be selected only one statement from the two types of the same dimension. For example, the introvert type is paired with the extrovert type, the thinking type is paired with the feeling type, the sensing type is paired with the intuition type, and the judging type is paired with the perceiving type. The paired comparison format is introduced by Thurstone (1927). This format is based on the forced-choice format (McDonald, 1999). It was the measurement model that forces the subject to response one of two or more statements.

The forced-choice is same as ranking format if there are two responses or statement that must be chosen. This format has two advantages to the response of a subject. It is avoided from social desirability and faking (McDonald, 1999, Barislow, 1958, Chernyshenko, 2009). Social desirability is a general statement in which people possibly respond to the statement dishonestly. It may be concluded that subject's response is known as faking. Therefore, it is difficult to detect the real opinion or response from the subject. Nevertheless, this type of instrument has weakness because the subject is forced to choose one statement though both of them are appropriate or inappropriate. On the other side, another model instrument, called the rating model, is responded freely. Subjects are free to respond to the statement as they wish though they may respond to it by faking. Thus, the weakness of forced-choice or ranking format can be covered by the rating format, and vice versa.

The comparison of the two formats of instrument has been studied by Setiawati (2013). She studied the scaling of multiple intelligence instrument using Thurstone types or ranking format and Likert type designed by a rating format. The results show that Likert-type instrument is more accurate than the Thurstone type is. The accuracy is caused by Likert type's variations of the answer which is more than the Thurstone type is. The Likert type has five responses while the Thurstone type has two responses. Thus, it needs further research to determine the accuracy of both types by creating the same variations.

The result of measurement using MBTI instrument is characteristic personality from several pattern of personality from many types. There are four dimensions of the two type of personality. Thus, the all patterns of personality have 16 alternate personality types. MBTI is widely used to determine differences in personalities in various fields, such as education, career development, organizational behavior, group functions, team development, personal and executive training, individual psychotherapy, couples, and families, and context of multicultural interaction (Center for Applications of Psychological Type, 2008). It design originally by ranking format, but it can be design by rating format. The Data of this study were collected using two format instrument.

This study aims to investigate the characteristics of the instrument MBTI and the types of personality of students in Yogyakarta by using the format type instrument with ranking and rating.

Method

The data were collected by giving questionnaire of MBTI instrument. It was modified into two formats, the forced-choice format (the original format) and the rating format. Both of format, were formed from the same construct and item. The specification items of the instrument can be showed in the table 1.

Table 1 Specification Items of the Instrument

Based on Preferences	Types of Preference	Number of Items	
		Forced-choice format	Rating format
Focusing direction	Extrovert (E)	1,9,19,29,35,41,53	1A,5A,10A,15A,18A,21A,27A
	Introvert (I)	2,10,20,30,36,42,54	1B,5B,10B,15B,18B,21B,27B
The way of obtaining information	Sensing (S)	3,13,17,31,39,45,55	2A,7A,9A,16A,20A,23A,28A
	Intuition (N)	4,14,18,32,40,46,56	2B,7B,9B,16B,20B,23B,28B
The way of making decisions	Thinking (T)	7,11,21,25,37,47,49	4A,6A,10A,13A,19A,24A,25A
	Feeling (F)	8,12,22,26,38,48,50	4B,6B,10B,13B,19B,24B,25B

			5B
The orientation towards the outside world	Judging (J)	5,15,23,27,33,43,51	3A,8A,12A,14A,17A,22A,26A
	Perceiving (P)	6,16,24,28,34,44,52	3B,8B,12B,14B,17B,22B,26B

From table 1, could be showed that both of two format have the same items but different in the number of item. Then, they were tested to subject of research. The subject of the research is undergraduate students of Yogyakarta State University. Meanwhile, to determine subject of research, the cluster random sampling was employing. They come from three study programs, the educational of math, the educational of economic, and guidance and counseling. The all subject were 134 undergraduate students.

Finally, there were two techniques to analyze the data: 1) analyzed the reliability instrument, 2) analyzed the characteristic of personality. The reliability of each dimensions instrument was analyzed using alpha Cronbach. Then, the all dimensions were analyzed by using composite reliability. The characteristic of personality subject was quantitative analyzed by descriptive statistic. All analysis of the data used Excel program.

Results

One of the characteristic instruments in classical theory was reliability. The concept of reliability in classical test theory assumptions is associated with observed scores (X), true score (T), and error score (E). The main concept of this theory is $X = T + E$, or the observed score is the combination of the true scores and error scores. Based on these assumptions, the concept of variant scores or total variance is the combination of the true score variance and the error score variance. The assumptions made in the formula $\sigma_t^2 = \sigma_r^2 + \sigma_e^2$ (Gulliksen, 1950, Lord & Novick, 1968, Allen & Yen, 1979, Thissen & Wainer, 2001).

The reliability of data instrument associated with measurement errors in the measurement data. Djemari Mardapi (2008) said that measurement error is part of unreliability data. It is studied in measurement of social sciences. Based on the concept of scores in the classical theory, the reliability of the measured data can be explained from the variant score. An association between the variant score as mentioned in the assumptions of classical theory can be used to explain the definition of reliability which is a variant of interaction errors and score. The concept of reliability can be formulated as $\rho_{xx'} = 1 - \sigma_e^2 / \sigma_x^2$,

where reliability is the magnitude of the error variance and variance score. Based on this formula, it can be explained that the larger the error variance is the smaller the coefficient reliability and will be vice versa. The score of reliability, variance score, variance error influence to the characteristic of instrument. The formula alpha from Cronbach was used to estimate the reliability from two formats. The result of data analyze could be seen in table 2.

Table 2. The Reliability and Standard Error Measurement (SEM) of Two Formats Instrument

No	Dimension	Ranking-format			Rating-format		
		Variance	Reliability	SEM	Variance	Reliability	SEM
1.	Extrovert	2,66	0,53	1,83	2,72	0,6	1,1
2.	Introvert	2,66	0,53	1,83	2,45	0,43	1,39
3.	Sensing	1,85	0,3	1,55	1,88	0,34	1,24
4.	Intuition	1,85	0,3	1,55	2,78	0,52	1,34
5.	Judging	3,09	0,6	1,94	2,71	0,5	1,34
6.	Perceiving	3,09	0,6	1,94	2,35	0,42	1,36
7.	Thinking	3,52	0,62	2,17	2,88	0,57	1,24
8.	Feeling	3,52	0,62	2,17	2,62	0,57	1,14
9.	Total instrument		0.74	1,87	31	0.72	1,27

Table 2 shows that the reliability instruments were various in many dimensions. The lowest reliability was sensing type and the highest reliability was thinking and feeling in force-choice type, and the extrovert type in the rating format. The lowest reliability was the sensing not only in the force-choice format but in the rating format as well. The all reliability instrument or composite reliability of ranking format was 0.74 and the rating format was 0.72. It showed that the reliability of forced-choice format was higher than rating format.

The data from forced-choice and rating format are different. Although the data from both types of instruments are different but there are connected. The calculation of coefficient correlation from many dimension or two formats are calculated using Pearson correlation. The result of correlation test shows that coefficient correlation or $r = 0.791$ and probability of significance or $p = 0.02$. Consequently, the two of formats are connected.

The term of error measurement related to the error variance or standard error of measurement (SEM). The formula of SEM or $\sigma_e = \sigma_t \sqrt{1 - \rho_{xx'}}$. SEM magnitude affects the reliability, it can be said that the accuracy of measurement results can be seen by the SEM. The smaller the SEM measurement will be more precise measurement results. Table 3 showed that the SEM of ranking format was higher than rating format, therefore, unless the ranking format have more reliable than the rating format, the rating format had more precise. It condition was influenced of the variance of ranking format was higher than the rating format.

The data from ranking and rating format are different. Although the data from both types of instruments are different but there are connected. The calculation of coefficient correlation from many dimension or two formats are calculated using Pearson correlation. The result of correlation test show that index correlation or $r = 0.791$ and probability of significance or $p = 0.02$. Consequently, the two of formats are connected.

The Profile of characteristic personality was analyzed by descriptive statistic using mode. The mode data from two format instrument could be seen in table 2.

Table 2 Descriptive Statistic the Data of Two Format Instrument

No	Dimension	Mode	
		Force-choice	Rating Formal
1	Extrovert	4	5
2	Introvert	3	5
3	Sensing	5	6
4	Intuition	2	4
5	Judging	5	6
6	Perceiving	2	5
7	Thinking	3	4
8	Feeling	4	5

Table 2 shows that the preferences of focusing direction, in force-choice format the extrovert is more dominant than introvert is, but in the rating format both types have similar score. Another score shows that to obtaining information, in the two formats instrument, the

sensing is more dominant than intuition is. Meanwhile, the way of making decisions, the feeling is more dominant than the thinking is in both of formats instrument, and the orientation towards the outside world, the judging is more dominant than perceiving is in the two formats. From the all preferences, the characteristic personality of subjects are extrovert, sensing, judging and feeling (ESJF). Consequently, the pattern of personality of Yogyakarta State University Students is ESJF.

The extrovert people have motivated and strong energy from outside and environment. The sensing people like taking information from their perception or sensation. The judging people have many plans and arrangements in their live. And the feeling people have many value in their lives and base on humanity in the interaction with the other people. From this characteristic, Yogyakarta undergraduate students prefer to outer world, prefer to do something ordered and scheduled, and strive for harmony with others and friendship.

The score of rating format is higher than ranking format is. It is coursed that the rating format might be to response high in the pairs of items. Choosing one statement out of several statements in ranking makes this type of instrument have difference characteristic to other instruments which provide a statement with various different responses. This instrument will be different responses for one stimulus. Hence, it will be obtained various kinds of responses on several provided stimuli. This is inline with Oliveres & Brown's (2010, p.935) stated that this type of instrument is spared from the same answers or there is bias in giving responses such as an extreme agreement response. It is also stated that by using this type of instrument, weaknesses in responding such as lack of various responses or 'halo effect' will be avoided.

The ranking instrument has strength related to subject response that tends to be spared from social desirability and faking (Mc Donald, 1999, p.24, Chernyshenko, et.al., 2009, p.108). Therefore, Social desirability is a general statement that tends to be responded by a subject dishonestly, that creates difficulties in finding out the real opinion of the subject. Meanwhile, *faking* is a tendency of subjects to choose responses that do not represent their characteristics or intentionally change their responses to make them included in the desired group. Thus, both formats instrument have weakness and strength in their object matter.

Conclusion

The findings of this study show that two formats of instrument are reliable in the both formats instrument using composite reliability. The reliability instruments were various in many dimensions. The lowest reliability was sensing type and the highest reliability was thinking and feeling in the ranking format, and the extrovert in the rating format. The lowest

reliability was the sensing not only in the ranking but in the rating format as well. The all reliability instrument or composite reliability of ranking format was 0.74 and the rating format was 0.72, It showed that the reliability of ranking format was higher than rating format.

The result of measurement personality characteristic of Yogyakarta undergraduate student is dominant extraversion in focus attention, dominant sensing in finding out about things, dominant feeling in making decision and dominant judging in orientation the other world. Consequently, Yogyakarta undergraduate students prefer to outer world, strive for harmony with others and friendship, and prefer to do something ordered and scheduled.

References

- Allen, M. J., & Yen, W.M. (1979). *Introduction to measurement theory*. Monterey : Brooks/Cole Publishing Company.
- Borislow, B. (1958). The Edwards personal preference schedule (EPPS) and fakability. *Journal of Applied Psychology*, 42,1, 22-27
- Boyd, R., & Brown, T. (2005). *Pilot study of Myers Briggs Type Indicator personality profiling in emergency department senior medical staff*. *Emergency Medicine Australia*, 17
- Briggs-Myers, I.B. & McCaulley, M.H. (1985). *Manual: A Guide to the Development and Use of The Myers-Briggs Type Indicator*. Palo Alto, CA: Consulting Psychologists Press.
- Chernyshenko, O. S., Stark, S., Prewett, M., Gray, A.A., Stilson, F. R., & Tuttle, M. D. (2009). Normative scoring of multidimensional pairwise preference personality scales using IRT: empirical comparisons with other formats. *Human Performance*, 22, 105–127.
- Djemari Mardapi. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendekian Offset.
- Gulliksen, H., (1950). *Theory of mental tests*. New York: John Wiley and Sons. Inc.
- Jung, C. (1971). *Psychological types* (H. G. Baynes, Trans., revision by R. F. C. Hull), *The Collected Works of C. G. Jung* (Vol. 6), Princeton, NJ: Princeton University Press, (Original work published 1921).
- Lord, F. M., & Novick, M. R. (1974). *Statistical theories of mental test scores*. New York: Addison-Wesley Publishing Company, Inc.
- McDonald R. P. (1999). *Test theory: a unified threament*. London: Lawrence Erlbaum Associates.

- McLellan, C. R. (2011). Differences in Myers-Briggs personality types among high school band, orchestra, and choir members. *Journal of Research in Music Education*, 59(1), 85-100.
- Olivares, A. M., & Brown, A. (2010). Item response modeling of paired comparison dan ranking data. *Multivariate Behavioral Research*, 45, 935-974.
- Thissen, & Weiner, D. H. (2001). *Test scoring*. Marwah: Lawrence Erlbaum Associated.
- Thurstone, L. L. (1927). Method of paired comparison for social values. *Journal of Abnormal and Social Psychology*, 21, 384-400.
- Setiawati, F. A. (2013). Perbandingan karakteristik psikometrik penskalaan tipe Likert dan Thurstone dengan teori klasik dan modern, studi pada instrumen MI. *Dissertation*. Yogyakarta State University.