

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.