ME - 53

THE CONSTRUCTION OF THE EARLY-CHILDHOOD TEACHERS' CREATIVE PROCESS ASSESSMENT

Risky Setiawan
(IKIP Veteran of Semarang, Indonesia)
Andika Farid Abdillah
(Edubright Institute, Indonesia)

Abstract

This study attempts to measure the early-childhood teachers' creativity which covers two main goals: 1) to construct an assessment model for pre-school teachers' creativity by Remote Associates Test and Torrance Test instrument; 2) to find out the Divergent Thinking and Creative Thinking skill level toward early-childhood teachers in Semarang. This study adopted the Research and Development which focuses on the measurement model. The study was conducted for one year period which was divided into three steps. The first step was Instrument and Drafting through forum group discussion, which involved eight experts (two measurement experts, three psychologist, and three pre-schools children experts). The second step was constructing testing which was intended to measure the instruments' validity and reliability. And the third step was measurement, which estimated the early-childhood teachers' creative abilities.

Keywords: assessment, creative, teacher, remotte associates test, torrance test

INTRODUCTION

The use of teacher's creativity assessment is important to be conducted since the early-childhood education, the most influential ability is child's creativity, because creative child can do anything optimally particularly in games activity indoor or outdoor.

Measurement is one program that must be done to find out someone's behavior and capability standard which covers some pre-determined measurements scales. Teacher's creativity in early-childhood education is often considered as potential and the second obligatory requirement compared to the teacher's professionalism. Belkhadas's theory (2010) said "Creative teaching to increase students' learning and achievement", creative learning will improve students' learning and knowledge. The study is very crucial to be conducted based on this assumption and theory.

CREATIVITY CONCEPTS

There are two popular definitions of creativity, the definition which refers or uses the expert's judgments or considerations and definition by criteria. The first definition is called consensual definition and the second definition (criteria consideration) is called conceptual definition.

Based on its emphasis (Amabile, 1983), creativity's definitions can be categorized into four types of different dimensions or Four P's of Creativity, they are person, process, press, and products. The person dimension of creativity as mentioned by Guilfors (1950): "*Creativity refers to the abilities that are characteristics of creative people.*" The definition of creativity

which emphasizes on the process dimension as proposed by Munandar (1977): "*Creativity is a process that manifest in self in fluency, in flexibility as well in originality of thinking.*" From the press dimension, Amabile (1983) argues that: "*Creativity can be regarded as the quality of product or response judged to be creative by appropriate observes*". And definition of creativity from the product dimension as stated by Baron (1976): "*Creativity is the ability to bring something new into existence.*"

Guilford with his factorial analysis found out that there were five traits which characterize thinking skills. First, fluency as the ability to produce ideas. Second, flexibility as the ability to propose kinds of approaches and/or solution toward problem/s. Third, the originality as the ability to produce authentic ideas as a result of their own thinking and not cliché. Fourth, elaboration as the ability to elaborate something in details. Fifth, the redefinition as the ability to review/re-evaluate a phenomenon based on different way and point of view to what is usually common.

Regarding the relationship between creativity and intelligence can be observed by studies conducted by psychologists. Torrance (1996) in his study found that children which possess high creativity have lower IQ in their peers. When we discuss about talent or giftedness, Torrance assumes, IQ cannot be used to the only measurement to identify gifted children. If we only use IQ to determine talent or giftedness, an estimated of 70% of children who possess high creativity will be eliminated.

Getzels and Jackson (1962) reported the finding in his study that there was no correlation between creativity and intelligence among students who get 120 in their IQ. It means that people with higher IQ might have lower level of creativity or vice versa. Based on the study we can assume that creativity and intelligence are two different domains in humans' ability in terms of nature and orientation. Within the context of correlation, intelligence cannot be used as the only criterion in identifying creative people.

THE INSTRUMENT DEVELOPMENT
1) Instrument Development

In Assessing Reading published by Cambridge University Press, Alderson (2000: 203) claimed that there is no the best testing method or assessment to read. The concept that must be understood is that the appropriate methodological choice for assessment because assessment which is carried out must have certain goals. There are some assessment techniques that we can use, they are close test or gap-filling test, multiple choice, matching technique, ordering test, short-answer test, free-recall test, the summary test, the gapped summary, information-transfer technique, real-life method.

a) Instrument Validity

The developed instrument in a study must have validity. In a traditional point of view, a test is classified as valid, if it can be used to measure anything that should be measured. At least there are four kinds of validity which is commonly used and considered as important in constructing instrument, they are predictive validity, concurrent validity, construct validity, and content validity.

In his description, Messick (Gipps, 1994: 59) focuses on social factors which has important position in an assessment. It is because assessment must be able to bring conformity, meaningful in depth, and useful. Based on the concept comprehension, Messick then describes it in a table two-by-two as follow:

Table 1. Messick's Facet Validity

	Test Interpretation	Test Use
Evidential Basis	*Construct Validity*	*Construct Validity + Relevance/utility*
Consequential Basis	*Value Implications*	*Social consequences*

b) Instrument Reliability

An instrument must meet the requirement of reliability. Reliability has something to do with test consistency or instrument which is used to measure what it supposes to measure. Gipps (1994: 67) said that Reliability is concerned with the accuracy with which the test measures the skill or attainment it is designed to measure. The underlying reliability question are: would an assessment produce the same or similar score on two occasions or given by two assessors? Reliability therefore relates to consistency of pupil performance and consistency in assessing that performance: which we may term replicability and comparability.

In Djemari Mardapi (2007: 18), there are ten steps which we have to follow in developing affective instrument, they are: 1) Determining the specification of instrument; 2) writing the instrument; 3) determining the scale of the instrument; 4) determining the scoring system; 5) analyzing the instrument; 6) testing; 7) analyzing the instruments; 8) composing instrument; 9) measuring; 10) interpreting the result of measurement.

A. Assessment to evaluate teachers

Assessment as stated in Measurement and Statistics for Teacher (Blerkom, 2009: 6) is a very general term that describes the many techniques that we have use to measure and judge students' behavior and performance. In relation to measurement and evaluation, assessment is an activity to assess students. Assessment is very important to support the goal of the curriculum target. There are three things that are related for outcomes assessment has tree stages by Rebecca (2009: 1), those are:

1. *defining the most important goals for students to achieve as a result of participating in an academic experience (outcomes)*
2. *evaluating how well students are actually achieving those goals (assessment)*
3. *using the results to improve the academic experience (closing the loop)*

There is a paradigm shift in assessment from psychometric model which now extends to the education assessment model, from test and cultural information which turned into cultural assessment itself (Gipps, 1994: 1). During its development, we now know the term of criterion-based assessment, formative assessment, performance assessment, alternative assessment, authentic assessment, and many more. However, conceptually there is a connecting line which becomes the core off all those assessments, that assessment must support the learning process rather than just as the indicators of the learning outcomes.

RESEARCH METHOD

The data analysis was done in multiplies methods. The first step used the quantitative analysis method which aimed to develop the instrument design (content validity) composed based on indicators which was constructed based on theories on Focus Gorup Discussion (FGD). Then, the researcher conducted the quantitative analysis to measure the instsrument design which was obtained from the test reponse of the content analysis (construct validity) supported by quantitative data aimed to see the construction which was done by the confirmatory analysis model.

Confirmatory analysis to assess the fit model criteria, namely: Chi Square and Probability, GFI (Goodness Fit Index), AGFI (Adjusted Goodness of Fit Index), and RMSEA (rootmean square error of aproximation), and Likelihood Estimation. The results of the CFA correlation of each indicator for each variable can be seen in the output pathdiagram. Analysis using SPSS version 17.0 and LISREL 8.0.

Then, the constructed instrument was tested to build an appropriate and fit instrument. Further analysis was done by estimating the teachers' ability in responding to the creative person and creative environment in which the Graded Response Model (GRM) and the Creative Product and Torrance Test instruments used the Inter-rater assessment which the instruments will be constructively tested with minimum of four observer and assessor. The estimated teachers' ability can be seen by using Item Response Theory by GRM (Graded Response Model). The estimated result of creative ability was used as comprehensive study and recommendations to Teachers Training Institutes and HIMPAUDI organization in Semarang.

RESULT AND DISCUSSION
I.   Validity and Reliability of the Instruments
A.   *Remotte Asociates Test* Instrument Validity
    a.   Exploratory Factor Analyses

Rotation was done by changing the factorial content pattern so we could obtain every single factor in every single variable. This study adapted the varimax method as the factorial analysis rotation. This method was able to produce factorial contents within the dominant variable. By doing the rotation several times then the item will be clustered into each factor. The factorial rotation can be seen in the table below:

Table 2. Factorial Content Matrix

	1	**2**	**3**	**4**	**5**
VAR00001	.776	-.047	.059	.043	.168
VAR00002	.763	-.142	.101	.145	.005
VAR00005	.690	.170	-.003	.132	-.122
VAR00003	.683	.071	.278	-.009	-.065
VAR00004	.534	.169	.094	.326	.037
VAR00010	-.026	.830	-.035	.115	.023
VAR00011	.093	.756	.069	.081	-.024

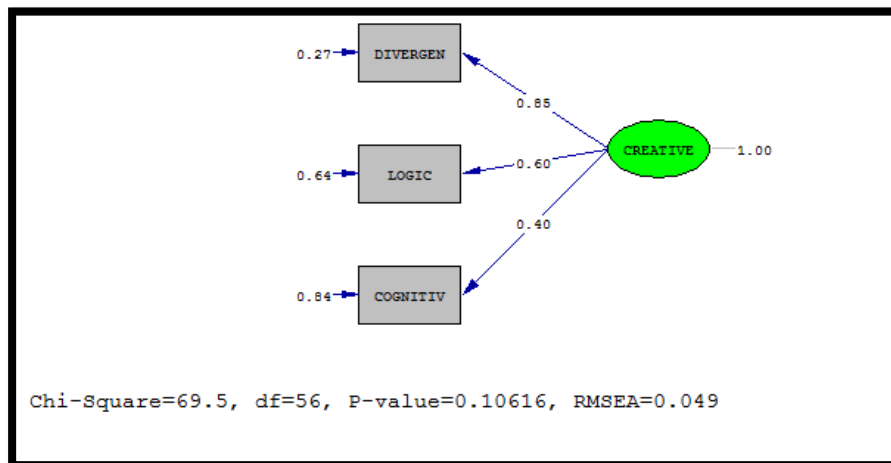|	1	2	3	4	5
VAR00013	.097	.671	.028	-.162	.145
VAR00006	.039	.663	-.268	-.034	-.123
VAR00012	-.216	.536	.265	-.140	-.048
VAR00008	.396	.635	-.058	.012	-.085
VAR00009	-.042	.548	.181	.302	.223
VAR00007	.320	.484	.123	-.018	.344
VAR00014	-.135	.471	.319	-.055	-.038
VAR00019	.259	.075	.649	-.190	.052
VAR00018	.372	-.089	.619	.025	.023
VAR00020	.193	.006	.598	.247	-.017
VAR00017	.122	.067	.580	.088	-.354
VAR00021	-.032	-.043	.512	.262	.194
VAR00022	.062	.104	.793	.139	-.420
VAR00023	-.017	-.062	.362	.792	-.101
VAR00025	.172	.024	.035	.790	.057
VAR00024	.162	-.104	-.010	.870	-.056
VAR00016	-.341	.056	-.148	.125	.280
VAR00015	-.234	.015	.138	.058	.443

The table shows the factorial content component matrix and the instrument's item forming factorial loading which makes up the factorial loading, showing correlation between variable 1 and factors 1, 2, 3, and so on. The process to determine which variable went to which factor was done by compering correlation among variables.

b.   Confirmatory Factor Analyses

The result of the confirmatory factor analysis could be used to obtain the latent variable data which was obtained from one free variable and eight latent variable consisting: motivation, achievement, infrastructure, perception, instructor, basic competence and activity which influenced the main variable, the Social Science teachers (X).

The next step, after the explanatory analysis was the confirmatory data reanalysis that had been categorized into 8 factors. The result of the confirmatory analysis showed that the most dominant factor influencing the teachers' performance was the motivation factor by PCA score more than 0.5 and by using the Goodness Fit Index (GFI) confirmatory analysis more than 0.4 and Comparative Fit Index (CFI) score between 0.18 to 0.31 and Rootmean Square Error of Approximation (RMSEA) which explained the residues within the model. Therefore, the expected value was very small, below 0.08. The result of the RMSEA designed model was 0,076 which proved that the teachers' performance assessment instrument model was closed fit.

The following is the path diagram which will describe the relationship among the indicator items between the latent and the main variables.

B.   RAT Instrument Reliability

  a.   Cronbach Alpha

The instrument reliability coefficient was basically calculated by the Alpha Cronbach formula. The level of instrument reliability was determined by the amount of the coefficient. The criterion used as the minimum reliability coefficient in this evaluation was 0.65. According to Mehrens, W.A & Lehman, L.J (1973: 122), if the level of reliability was the same or more than 0.65, so we could say that the instrument was quite good. Djemari Mardapi (2008: 121-122) stated that if the instrument had been analyzed, then evaluated and were arranged for testing. The testing aimed to determine the instrument characteristics. The most important characteristics was the discriminating power and the reliability power. The higher of the response variation, the better the instrument. When a variation of an item was small, we could conclude that the item was not a good variable.

Table 3. The Test Item Alpha Cronbach Coefficient Review

No.	Instrument /Respondent	Coefficient *Alpha*	Comparison	Result
1	Deduksi	0.798	≤ 0.65	Reliable
2	Logika Kognitif	0,670	≤ 0.65	Reliable
3	Logika Gambar	0,830	≤ 0.65	Reliable
	Total Cronbach	0,746	≤ 0.65	Reliable

The table shows the result of the instrument reliability in which higher score, more than 0.65 indicates reliable test item.

C.   Torrance Test Reliability Instrument

The observation instrument (ratings) is the scoring procedure based on subjective judgment toward particular aspects or attributes, which is done through indirect or indirect systematic observation (Syaifudin Azwar, 1992: 105). To reduce the subjective scoring, so it was conducted by more than one person (rater).

Rating was done by some different and independent rater toward the same group of subjects. Although there might be possibility of errors, but the variant errors could be minimized compared to the re-rating procedure by a single rater only. Ratings done by many people will emphasizes on the inter-raters reliability. Ebel (1951) in Syaifudding Azwar (1992: 106), mentioned the formula to estimate the reliability of the rating result conducted by as many as *k* raters to as many as *n* subjects.

The following formula was used to find the average inter-correlation coefficient of the rating result of all rater pair's combination that was made and was indeed the average reliability of a rater.

$$r_{xx'} = \frac{S_{S^2} - S_{e^2}}{S_{S^2} + (k-1)S_{e^2}}$$

where,
S_{S^2} = inter-subjects variance that is rated
S_{e^2} = error variance: variance interaction between subjects (s) and raters   (r)
k = number of rater giving the rating

In addition to the estimated formula, there was also another formula that could be used to estimate the average reliability done by *k* raters.

$$r_{xx'} = \frac{\left(S_{S^2-S_{e^2}}\right)}{S_{S^2}}$$

where,

S_{S^2} = inter-subjects variance that is rated

S_{e^2} = variance error, interaction variance between subjects (s) and raters (r)

To calculate S_{S^2} and S_{e^2}, the researcher used the following formulas:

$$S_{e^2} = \frac{\sum_i 2 - \frac{(\sum_R 2)}{n} - \frac{(\sum_T 2)}{k} + \frac{\sum_i 2}{nk}}{(n-i)(k-i)}$$

$$S_{S^2} = \frac{\frac{(\sum_T 2)}{k} - \frac{(\sum_T 2)}{nk}}{n-1}$$

where,

i   =   the rating score given by a rater to a subject

T   =   the total of rating score received by a subject from all raters

R   =   the total of rating score given by a rater to all subjects

N   =   number of subject

K   =   number of rater

The result of the inter-rater reliability on the observation instrument can be explained as follow:

n = 100 participants

k = 3 raters

$$S_{e^2} = \frac{\sum_i 2 - \frac{(\sum_R 2)}{n} - \frac{(\sum_T 2)}{k} + \frac{\sum_i 2}{nk}}{(n-i)(k-i)}$$

$$= \frac{1867363 - \frac{184159613}{101} - \frac{5564953}{3} + \frac{23459}{101x3}}{(100-1)(3-1)}$$

$$= \frac{5269,23432}{198}$$

$$= 26,3461716$$

$$S_{S^2} = \frac{\frac{(\sum_T 2)}{k} - \frac{(\sum_T 2)}{nk}}{n-1}$$

$$= \frac{\frac{5564953}{3} - \frac{23459^2}{101x3}}{100-1}$$

$$= \frac{38731,2607}{99}$$

$$= 387,312607$$

$$r_{xx'} = \frac{\left(S_{S^2-S_{e^2}}\right)}{S_{S^2}}$$

$$= \frac{387,312607 - 26,3461716}{387,312607}$$

$$= 0,931977$$

Thus, the reliability of the average rating of the three raters was 0.93 ($\geq$ 0.65), so the researcher concluded that the average reliability index was consistent.

$$r_{xx'} = \frac{S_s^2 - S_e^2}{S_s^2 + (k-1)S_e^2}$$

$$= \frac{387,312607 - 26,3461716}{387,312607 + (3-1)26,3461717}$$

$$= \frac{360,9664}{387,312607 + 52,69234}$$

$$= 0,820369$$

While, the estimated reliability average to a rater was 0.82. So, the consistency of a rater was considered as good. So, the conclusion was that the Torrance Test Instrument was reliable and good.

CONCLUSION

1. The level of the Remote Associate Instrument (RAT) validity and realibility was sufficient enough based on the content validity criterion.
2. The level of Torrance Test (TT) Instrument Reliability was sufficient enough based on the inter-rater testing criterion.
3. The Item Response Theory Analysis applied showed that 14 items was fit to the 2PL model, while the information function had the highest precision and the Standard Error Measurement had the lowest precision to the 2PL.

REFERENCES

Amabile Theresa. (2012). *Componential Theory of Creativity*. Harvard Bussiness School.

Barron, F. (1969). *Creative person and creative process*. New York: Holt.

Belkhadas. (2010). A Dessertation of Doctoral: Creative Teaching to Increase Students' Learning and Achievement the Case of English Teachers - University of Constantine. Published

Cartwirght Rebecca. Revised August, 2009. Student Learning Outcomes Assessment Handbook., Montgomery County, Maryland: Montgomery College

Djemari Mardapi. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendikia Jogjakarta.

Getzels, J. W. & Jackson, P. J. (1962). *Creativity and Intelligence: Explorations with Gifted Students*. New York: John Wiley and Sons, Inc.

Gipps CV. 1994. Beyond testing: Towards a theory of Educational assesment. Falmer Press. London.

Guilford, J.P., (1977), *Way Beyond the IQ, Buffalo, Creative Learning Press*. Reni Akbar dkk, (2001), Kreativitas, Panduan bagi Penyelenggaraan Program Percepatan Belajar, Grasindo, Jakarta.

Torrance, E.P.(1976) *Future Careers for Gifted and Talented Students Gifted Child*, Quarterly 20: 142-156.