CONSTRUCT VALIDITY OF CREATIVE THINKING SKILLS INSTRUMENT FOR BIOLOGY STUDENT TEACHERS IN THE SUBJECT OF HUMAN PHYSIOLOGY

Andi Ulfa Tenri Pada [1] *, Siti Salina Mustakim [2], Bambang Subali [3]
[1]Universitas Syiah Kuala, [2]Universiti Putra Malaysia, [3]Universitas Negeri Yogyakarta
[1]Jl. Teuku Nyak Arief No.441, Kopelma Darussalam, Syiah Kuala, Kota Banda Aceh, 24415
[2]43400 Seri Kembangan, Selangor, Malaysia
[3]Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia
* Corresponding Author. Email: andi_ulfa@unsyiah.ac.id

Abstract

This article discusses the construct validity of the creative thinking skill instrument supporting a conation idea aspect for biology student-teachers in the subject of Human Physiology. Two hundred and eighteen students participated in this study. Construct validity was obtained through Confirmatory Factor Analysis technique (CFA). Reliability was estimated by composite reliability. Findings show that the construct validity and the reliability of the instrument are high. A thorough discussion on the findings and future implications are provided towards the end of this paper.

Keywords: *construct validity, creative thinking skill, instrument development*

Introduction

Future learning challenges demand the development of creative thinking skills as one of the issues for 21st century literacy (Widowati, 2009). Development in information and communication technology requires high-order thinking competencies that include creative thinking skills, to be developed through evaluation processes. It is unfortunately true that, up to now, learners' thinking skills have not been seriously dealt with.

In the study by Rofi'uddin (2000), discontents are found concerning the low levels critical-creative thinking competencies in the graduates of elementary to tertiary levels. This research finding is in line with the results of international creativity survey highlighted in The Global Creativity Index conducted in 2011. The study indicated that Indonesia was at the rank of 81st from the 82nd countries in the survey, far below neighbouring countries Singapore (Rank 9) and Malaysia (Rank 48) (Florida, Mellander, & Stolarick, 2011). Meanwhile, according to the Human Development Report in 2013, Indonesia's Human Development Index (HDI) was at the rank of 110, while Singapore 11 and Brunei Darussalam 31 respectively. These two neighbouring countries, Singapore and Brunei Darussalam, had a very high HDI. Indonesia was underscored far left behind. Malaysia (HDI 62) was at the high category. In spite of the fact that Indonesia was in the same group as Thailand, at the medium category of HDI, the later was at rank 93, far above Indonesia.

In lieu toward the development of Higher Order Thinking Skills (HOTS), an international survey shows that Indonesian human performances are still minimal. A study by Ramirez & Ganaden's (2008) reveals that poor human performances are caused by incapability of infusing the high-order thinking skills. Learning in higher education seems to be not as effective as to develop creative thinking. Meanwhile, creative thinking is listed as the top skill in the cognitive domain of Bloom's taxonomy as revised by Anderson (Krathwohl, 2002) and the *New Blooms* (Dettmer, 2006). Furthermore, DeHaan's (2009) study shows that creative thinking competencies help learners to find evidence-based analyses, improve HOTS and solve problems. Thus, creative thinking skills need to be further developed and measured.

With respect to the needs of creative thinking skills, it is believed that learners of Biology with HOTS will be able to think creatively and solve problems effectively in completing project tasks provided such as determining objects to observe, gathering information from relevant sources, and using various ideas in and outside the campus. Problem-solving competencies will also be useful for students after they become teachers, who will encounter situations with problems to be solved. This statement goes in line with a study conducted by Husain, Mustapha, Malik, & Mustakim (2014) which highlighted how interaction between environment and student influences student's development in learning. These competencies include developing instructional contents, providing effective classroom management, assessment, and preparing the teaching and learning for 21st century.

In the present time, the measurement of creative thinking is conducted in a general manner. One divergent thinking test that is widely known is the Torrance Tests of Creative Thinking (TTCT). Torrance takes creativity as independent cognitive competence and a multidimensional concept. In contrast, Weiping Hu and Philip Adey maintain that creativity is part of intelligence and an independent skill. They developed a scientific creativity test for high-school students (Hu & Adey, 2002). The test, known as A Scientific Creativity Structure Model (SCSM), was built on the analyses of meaning and a scientific creativity aspect based on literary studies. Hu and Adey did an analysis on general themes from science and creativity and built a three-dimensional model from the characteristics of creativity such as features, processes, and products. Fluency, flexibility, and originality were features in SSCM. Scientific products were related to the scientific field such as technical

products, scientific knowledge, scientific phenomena, and scientific problems. They tried to include all their scientific creativity model components in a framework and test. For instance, they included scientific thinking processes and imagination, which often were not asked in a creativity test. Based on these reviews of the measurement of creative thinking, it can be seen that there is a wide opportunity to have creative thinking skills integrated in the school lessons. Studies focusing on the development of creative thinking skill evaluation in school lessons have been conducted (Beghetto, 2013). These studies show that creative performance is moderately a specific domain and can be appraised by a combination of particular power resources (Sternberg, 2006). This is supported by Kaufman and Baer Charyton, Ivcevic, Plucker, & Kaufman (2009) and Barbot, Besançon, & Lubart (2011) that creativity tends to be a specific domain. One may certainly be creative in one domain and a number of persons may be creative in two or more domains. One may most possibly take risks in their own domain for they have a higher level of content.

The choice on the subject of Human Physiology is based on the consideration that, in this subject, students' conative idea responses can be obtained more easily seeing that the analysed cases are contextual. When studying human physiology, students learn about the normal functions of the body organs; therefore, in the instrument, the given stimuli are disrupters to body organs (diseases) or the opposites of the normal functions of the physical organs. The expectation is that, by giving abnormal phenomena as stimuli, learners are able to give various responses in the forms of solutions to problems that are presented in various cases.

Evaluation of learners' creative thinking can be done in an evaluation system of the divergent pattern (Subali, 2011, pp. 130–144). Divergent thinking is a skill in constructing or producing various possible responses, ideas, options, or alternatives against a problem (Isaksen, Dorval, & Treffinger, 1994, p. 18). In other words, divergent thinking can be understood as a skill in producing various solutions to a problem using the correct procedure and reasons.

Creating thinking competence is one of the important thinking skills so that it must be mastered by the students. Moreover, the competency-based curriculum in the university brings heavy emphasis on the practice of thinking and reasoning, developing creative activities, problem-solving skills, and communicating ideas.

The evaluation kit to measure creative thinking skills that supports the ideas of teacher-candidate students of biology through a divergent pattern is expected to become one of the ways to improve students' creating thinking skills. The students' logical thinking will be directed to producing arguments based on their concept of understanding in the form of conative ideas. As a result, teachers will be able to see students' divergent production patterns in the form of rational alternatives towards given stimuli so that they can explain concepts that are contradictive to the actions.

The evaluation instrument must be tested for its quality. Quality testing is one way to show that an evaluation instrument has been optimally developed. The primary evidence of the quality of an evaluation instrument is its validity. Messick defines validity as "one that is integrative on how far empirical evidences and theoretical rationales support the feasibilities of interpretations and actions based on the results of the measurement processes" (Reynold, Livingstone, & Wilson, 2009). In line with Messick, the Standards for Educational and Psychological Testing defines validity as "how far evidences and theories support the interpretation of tests scores as a consequence of the using of the test" (American Educational Research; Association American Psychological Association; & National Council on Measurement in Education, 1999).

The validity measure that will be discussed in this article is of the construct type (construct validity), i.e. the construct of the instrument from theories supported by field data (Messick, 1989; Silverlake, 1999). Con-

struct validity is understood as how far the sores of the results of the measurement reflect the theoretical constructs underlying the development of the instrument (Suryabrata, 2000). Hadi (2001) states that construct validity is identical with logical validity or validity by definition. Hadi (2001) also states that, if the theoretical construct of a test is right, then the product of the measurement is valid.

Empirical testing is needed in order to find how far each variable to be measured can be explained by each dimension in the instrument. An instrument is regarded as qualified after it is analysed theoretically and empirically. Theoretical analysis is done by way of instrument review. Empirical analysis is done by way of factory analysis.

Factor analysis techniques come in two types, namely Exploratory Factor Analysis (EFA) and Confirmatory factor analysis (CFA). EFA is used for exploring the theories; CFA for confirming the theories that are obtained in EFA. CFA is a popular statistical technique for providing supports to construct validity in the literature of psychological testing (American Educational Research; Association American Psychological Association; & National Council on Measurement in Education, 1999; Thompson & Daniel, 1996). The present article shows the utilization of EFA and CFA to provide evidence for construct validity in the evaluation instrument development. In particular, the study is aimed at giving evidence to the construct validity of the creative thinking test instrument for teacher-candidate students of biology. A review of CFA, based on theoretical evidence, is an important part of the validation process (DiStefano & Hess, 2005, p. 228).

Method

The study was a confirmatory research which involved 218 students of Biology Programme in the Faculty of Teacher Education. Respondents were selected based on their prior knowledge on Human Psysiology learning environment. Implementation of the practical guidelines for the factor analysis technique followed the ones given by Cattell (1978) dan Guilford (1954) suggesting N > 200 as the number for sample size in the study.

The study used instruments validated by a panel of Subject-Matter Experts (SMEs) in the fields of Educational Assessment, Educational Evaluation, Biology Education, and Human Physiology for their contents (Hammitt & Zhang, 2013). The quality of instruments was also observed through their empirical validity and reliability. The instrument validity was appraised by internal consistency of construct indicators to show the degree to which each indicator indicated a general latent construct.

The data analysis began with EFA, which was preceded by Kaiser Meyer Olkin Meisure (KMO) testing and Bartlett's Test on the SPSS, and proceeded with CFA using the Lisrel software. According to Hair, Black, Babin, & Anderson (2006), evaluation on the levels of suitability between the data and the model is done through a couple of fits, i.e. the overall model fit and the measurement model fit.

Findings and Discussion

Findings

Data analysis was made first by conducting factor analysis which was preceded by KMO testing and Bartlett's Test in order to determine the suffifiency of the sample. The result showed a KMO MSA value of $0.821 > 0.05$ with a p-value $0.000 < 0.05$. This finding showed that the data matrix had sufficient correlation measure to be used to conduct a factor analysis. Subsequently, analysis from the communality values showed that not all items had a value higher than 0.30, the acceptable minimum value (Mooi & Sarstedt, 2011). From the EFA, it was found that 10 factors had an eigen value hogher than 1, and four had an eigen value higher than 2. In this study, an eigen value higher than 2 is preferred.

Results of the EFA analyses showed that the instrument consisted of 37 test items distributed in four factors. The four factors

were: (1) solution alternative, the skill to produce a number of solutions to respond to a problem formulated in 10 test items; (2) original solution, the skill to produce a number of solutions that are relevant and unique, consisting of eight test items; (3) solution feasibility, the skill to produce a number of solutions effective for the solution of the problem in the given case, consisting of 10 test items; and (4) solution variety, the skill to produce a number of solution categories, consisting of nine test items.
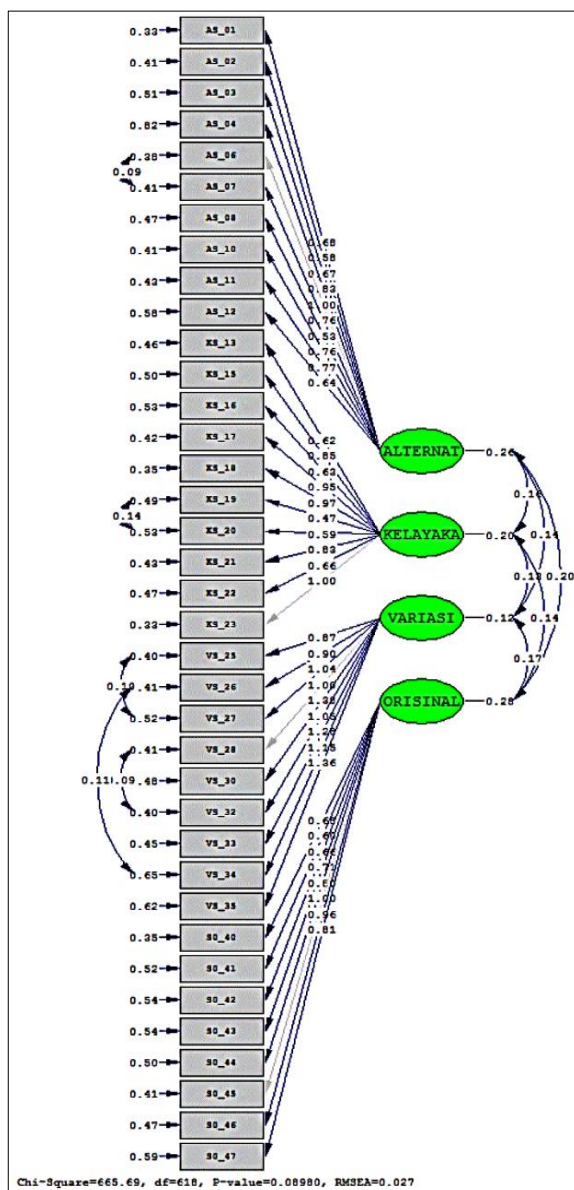
Overall Model Fit

Figure 1. Model of the Creative Thinking Skill Measurement that Supported the Conative Aspect

The first phase of the model-fit testing was directed to evaluating the general degree of the goodness of fit (GOF) between the data and the model. Some comparative- fit indexes emerged that differentiated one model from the base model. Practical guidelines were used to determine the accepted degree of fits since the sampling distribution of the indexes was unknown (Shook, Ketchen, Hult, & Kacmar, 2004). Hence, until a definite measurement was developed should researchers used some steps to provide evidence for their model fits (Breckler, 1990). The use of several indexes would ensure the readers that researchers did not choose only the favourable indexes. The study by Gerbing & Anderson (1992) shows that, among the strong and stable indexes, the normed fit index (NFI), and the comparative index (CFI) are two of these (Hu & Bentler, 1999). Results of the data analysis using the CFA technique can be seen in Figure 1.

In Figure 1, the Chi Square obtained = 665,69, *df* = 618, *p.* value = 0.08980 and RMSEA = 0,027. This indicates that the model has a fit. This is in line with the criteria as shown in Table 1.

Measurement Model Fit

In obtaining the conclusion that the fit between the data and the model was in general good, the next step was the evaluation of the measurement model fit. This was done by looking at the validity and reliability of the measurement model (Hair et al., 2006). These are presented as follows.

Convergent validity is used to test construct validity. The word construct refers to a theoretical view to explain some phenomena (Wiersma, 2000). According to Van Dalen (1973), a construct customarily refers to a complex concept that covers a number of inter-related factors.

Table 1.   Results of Overall Model Fit in the Initial Try-Out

GOF Measure	Degree Target of Fit	Estimation Results	Degree of Fit
Normed χ^2	Normed $\chi^2 < 2$	1.07	Good (*good fit*)
p-value	$p \geq 0.05$	0.089	Good (*good fit*)
RMSEA	RMSEA ≤ 0.05	0.027	Good (*close fit*)
ECVI	Smaller values & close to ECVI saturated	M* = 7.67 S* = 12.90 I* = 30.27	Good (*good fit*)
AIC	Smaller values & close to AIC saturated	M* = 835.69 S* = 1406.00 I* = 3436.78	Good (*good fit*)
NFI	NFI ≥ 0.90	0.81	Not so Good (*marginal fit*)
GFI	GFI ≥ 0.90	0.75	Not so Good (*marginal fit*)
CFI	CFI ≥ 0.92	1.00	Perfect (*perfect fit*)
RMR	SRMR ≤ 0.09	0.079	Good (*good fit*)

M = Model; S = Saturated; I* = Independence

An indicator is said to have good validity towards the construct or latent variable if (a) the *t*-value of its factor loading is higher than the critical value (*t*-value $\geq$ 1.96) (Doll, Xia, & Torkzadeh, 1994; Hair et al., 2006); and (b) the standardized factor loading is $\geq$ 0.30 (Gorsuch, 1983; Mooi & Sarstedt, 2011). The *t*-value and the standardized factor loading of the creative thinking skill measurement model are presented in Table 2.

In Table 2, it can be seen that all the *t*-values of the factor loadings of the variables or items are higher than 2 (*t*-value > 2). The factor loadings of the variables in the model are therefore are significant, or are not zero. Each of the standardized factor loading of each variable is higher than the minimum value (standardized factor loadings > 0.3). Thus, it can be concluded that the validity of all the variables observed towards the latent variable is good.

Evaluation on the measurement model can be done by using the composite reliability measure, or also known as construct reliability (CR) (composite reliability) (Ghadi, Bakar, Alwi, & Talib, 2012; Hair et al., 2006). A construct can have a good reliability measure when its value is the same or higher than 0.70 (Hair et al., 2006; Lance,

Butts, & Michels, 2006). CR can be calculated by the total of the squared factor loading (Li) factor loading added with total error variance of a construct (e$_i$).

$$CR = \frac{(\sum_{i=1}^{n} Li)^2}{(\sum_{i=1}^{n} Li)^2 + (\sum_{i=1}^{n} ei)}$$

From the results of the computation of the construct reliabilities in Table 2, it can be concluded that the construct reliability of the instrument model is good (CR $\geq$ 0.70). CFA is done to estimate the factor loading of a variable. Factor loading presents the level of the regression path from the latent to its indicator. The CR level is an alternative guide to review convergent validity (Ghadi et al., 2012).

The model being tested consisted of four latent variables and 37 observed variables. The model, supported theoretically and empirically, placed four factors separated but correlated. The factors solution alternative consisted of 10 items, original solution eight, solution feasibility 10, and solution variety nine. CFA results showed a good model fit for the instrument with a sample of 370 students. In addition, the moderate correlation among the four factors (SLF $\geq$ 0.30) showed that these factors considered the constructs

different but having relations (Mezo & Short, 2012).

Table 2. T-Value, Standardized Factor Loadings & Construct Reliability

Item	SLF*	t-Value	CR
Solution Alternative Factor			
AS_01	0.52	4.44	0.74
AS_02	0.42	3.72	
AS_03	0.43	3.81	
AS_04	0.42	3.74	
AS_06	0.64	**	
AS_07	0.52	5.11	
AS_08	0.37	3.29	
AS_10	0.52	4.46	
AS_11	0.52	4.45	
AS_12	0.4	3.52	
Solution Feasibility Factor			
KS_13	0.38	3.31	0.71
KS_15	0.47	4.01	
KS_16	0.36	3.16	
KS_17	0.55	4.53	
KS_18	0.59	4.78	
KS_19	0.29	2.58	
KS_20	0.34	3.01	
KS_21	0.49	4.14	
KS_22	0.39	3.43	
KS_23	0.61	**	
Solution Variety Factor			
VS_25	0.43	3.49	0.73
VS_26	0.43	3.52	
VS_27	0.44	3.57	
VS_28	0.47	**	
VS_30	0.56	4.14	
VS_32	0.5	4.37	
VS_33	0.55	4.08	
VS_34	0.44	3.56	
VS_35	0.51	3.91	
Solution Originality Factor			
S0_40	0.52	4.63	0.73
S0_41	0.41	3.7	
S0_42	0.43	3.91	
S0_43	0.46	4.11	
S0_44	0.51	4.57	
S0_45	0.64	**	
S0_46	0.6	5.17	
S0_47	0.49	4.37	

*SLF = *Standardized loading factor*
** = *Reference variable*, *t*-value is not estimated. *t*-value target ≥ 2.

Discussion

The research results show that the instrument for testing creative thinking skills developed for biology teacher-candidate students has a high construct validity. This finding is in agreement with Batey (2012) who stated that creativity has a relation with learning and education, more particularly in the matter of problem solving. This can be combined with the emphasis in the contextual approach that creativity is a social phenomenon underlying the interaction between the individual and the situation (Barbot et al., 2011; DeHaan, 2009). Measurement of learning achievement at the higher-order cognitives, such as creative thinking, demands tasks that require learners to use their knowledge and skills in a new situation (Nitko & Brookhart, 2007). Therefore, learners are demanded not only to be able to understand, but also to be able to analyse, evaluate, and create.

The measurement has an emphasis more on tasks/problems that are oriented to the real world than the school contexts. In line with this, Guilford (1954) stated that stimuli in the forms of conditions/situations may change learners' behaviours. Guilford further stated that creative actions are an example of the outcomes of a learning process that shows changes in behaviours resulting from stimulation and responses. The present study shows that the values of the standardized loading factors are all above 0.30 and the *t*-values of the observed variables are ≥ 2.00 with a confidence level of 95%. These show that the validity of all the observed variables against the latent variables is good. In other words, all the items are valid for testing the creative thinking skills in the Human Physiology subject.

Scientific creativity among 130 students was studied in Taiwan by Liang (2002). The primary instruments included Test of Divergent Thinking (TDT) for measuring creativity, Creativity Rating Scale (CRS) and Creative Activities and Accomplishments Check Lis-ts (CAACL) for measuring scientific creativity, Nature of Scientific Knowledge Scale (NSKS) for measuring the nature of science, and Science Attitude Inventory II (SAI II) for measuring attitudes towards science. There were in addition two more

instruments for measuring learners' competences in finding problems and formulating hypotheses. Data analysis techniques included descriptive statistics, Pearson product-moment correlations, and gradual regression. The findings show that, among others, students' scientific creativity is significantly correlated with attitudes towards science, finding problems, formulating hypotheses, nature of science, resistance towards closure, originality, and elaboration. Another finding is related to the emergence of four predictors that are significant for measuring scientific creativity; namely attitudes towards science, finding problems, resistance towards closure, and originality that contribute 48% to the variance of students' scientific creativity. Still another finding is concerned with the big difference between students with high scientific creativity from those with low scientific creativity on the variables family support, career picture, and reading about science. It can be concluded that both cognitive and non-cognitive components are good for predicting scientific creativity.

In the meantime, Barbot et al. (2011) conducted a study to give a real picture and a different way to measure creative potentials that can be used for educational objectives. The study produced a model for evaluating creativity called EPoC (Evaluation of Potential for Creativity) that was multifaceted and domain-specific, and that made it possible for the evaluator to capture the multidimensionality of creative potentials in obtaining potential profiles for creativity. It was a procedure for measuring students' creative potentials by way of multivariate approaches using verbal and graphic tasks for two creative thinking competences: *divergent thinking* (DT) and *integrative thinking* (IT). The creativity phases were designed in two sessions, within each of which each thinking competence was measured. In the DT assignments, each student was asked to produce as many ideas in responding to a unique stimulus. In the IT assignments, students were asked to produce synthetic solution elaborations. The creativity score for each task was based on the total of outputs

(for the DT) and level of originality. The higher score, the higher the potentials of the individual's creativity.

A study in the measurement of creativity in the field of science was done by DeHaan (2009) exploring the relation between creativity and high order cognitive skills (HOCS), reviewing evaluation techniques, and describing learning strategies to improve creative problem solving in the university. DeHaan used Torrence's creative thinking test that dealt with students' divergent thinking competences to measure their scientific creative thinking. In his opinion, creativity is not a thing difficult to measure. Creative processes can be explained by using cognitive competences which are better known such as cognitive flexibility and resistance control that have been widely spread in the society. Creativity is an important element of problem solving and critical thinking. Consequently, creativity application such as creative power and intelligence are components of the HOCS as defined in the taxonomy of educational objectives. It is no wonder that creativity, as other elements in HOCS, can be taught effectively through inquiry-based instruction founded by the constructivist theory.

Finally, reliability of the instrument was estimated by the technique of composite reliability measure or also known as construct reliability (Hair et al., 2006). A high construct reliability shows internal consistency, meaning that all the steps in the measurement consistently represent the same latent constructs. The results of the reliability calculation in Table 2 show that all the reliability values of the constructs are ≥ 0.7; thus, it can be concluded that the reliability of the measurement is categorically good.

Conclusion

Based on the research results and discussion, a number of conclusions can be proposed. First, the instrument for measuring the creative thinking skills of biology teacher-candidate students possesses a high degree of construct validity. Second, the instrument is built of four factors namely: (1)

solution alternative, the skill to produce a number of solutions in responding to a problem, consisting of 10 test items; (2) original solution, the skill to produce a number of solutions that are relevant and unique, consisting of eight test items; (3) solution feasibility, the skill to produce a number of solutions that are effective to solve the given problem, consisting of 10 test items; and (4) solution variety, the skill to produce a number of categorical solutions, consisting of nine test items. Third, the instrument consists of 37 items; each having a high loading factor against the latent variable. Fourth, besides having a high construct validity, the instrument is also characterized with a high composite reliability.

Implication from the results of the research study is that the instrument for measuring the creative thinking skills of biology teacher-candidate students is feasible to be used. In relation with the research results, those who need to measure students' creative thinking skills are recommended to use this instrument bearing in mind that the respondents have identical characteristics as the ones in the present study.

References

American Educational Research; Association American Psychological Association; & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.

Barbot, B., Besançon, M., & Lubart, T. I. (2011). Assessing creativity in the classroom. *The Open Education Journal, 4*(1), 58–66.

Batey, M. (2012). The measurement of creativity: from definitional consensus to the introduction of a new heuristic framework. *Creativity Research Journal, 24*(1), 55–65.

Beghetto, R. A. (2013). Nurturing creativity in the micro-moments of the classroom. In K. H. Kim, J. C. Kaufman, J. Baer, & B. Sriraman (Eds.), *Creatively Gifted Students are not like Other Gifted Students: Research, Theory, and Practice*. Sense Publishers.

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: cause for concern. *Psychological Bulletin, 107*(2), 260–273.

Cattell, R. B. (1978). *The scientific use of factor analysis*. New York: Plenum.

Charyton, C., Ivcevic, Z., Plucker, J. A., & Kaufman, J. C. (2009). Creativity assessment for evaluating the effectiveness of educational programs. In C. Schreiner (Ed.), *Handbook of Research on Assessment Technologies, Methods, and Applications in Higher Education* (pp. 78–96). Hershey: IGI Global.

DeHaan, R. L. (2009). Teaching creativity and inventive problem solving in science. *Life Sciences Education, 8*, 172–181.

Dettmer, P. (2006). New Blooms in established fields: Four domains of learning and doing. *Roeper Review, 28*(2), 70–78.

DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment, 23*(3), 225–241.

Doll, W. J., Xia, W., & Torkzadeh, G. (1994). A confirmatory factor analysis of the end-user computing satisfaction instrument. *Management Information Systems Quarterly, 18*(4), 453–461.

Florida, R., Mellander, C., & Stolarick, K. (2011). *Creativity and prosperity: The global creativity index*. Toronto, Ontario, Kanada: Martin Prosperity Institute.

Gerbing, D. A., & Anderson, J. C. (1992). Monte Carlo evaluations of goodness of fit indices for structural equation models. *Sociological Methods and Research, 21*(2), 132–160.

Ghadi, I. N., Bakar, K. A., Alwi, N. H., &

Talib, O. (2012). Construct validity examination of critical thinking dispositions for undergraduate students in University Putra Malaysia. *Higher Education Studies*, *2*(2), 138–145.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale: Erlbaum.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.

Hadi, S. (2001). *Metodologi research (Jilid III)*. Yogyakarta: Andi Offset.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2006). *Multivariate data analysis* (7th ed.). Prentice Hall.

Hammitt, J. K., & Zhang, Y. (2013). Combining experts' judgments: comparison of algorithmic methods. *Risk Analysis*, *33*(1), 109–120.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55.

Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science & Education*, *24*(4), 389–403.

Husain, M. Y., Mustapha, R., Malik, S., & Mustakim, S. S. (2014). Validation of instruments learning environment: exploratory factor analysis (EFA). In *10th Conference of AASVET (Asian Academic Society for Vocational Education and Training)*. Tokyo.

Isaksen, S. G., Dorval, K. B., & Treffinger, D. J. (1994). *Creative approaches to problem solving*. Dubuque: Kendall/Hun.

Krathwohl, D. R. (2002). *A revision of Bloom's taxonomy : an overview. theory into practice*. Taylor & Francis, Ltd.

Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria what did they really say? *Organizational Research Methods*, *9*(2), 202–220.

Liang, J.-C. (2002). *Exploring scientific creativity of eleventh grade students in Taiwan*. The University of Texas at Austin. Retrieved from https://pdfs.semanticscholar.org/470a/a97e1cd2c757315b9f9330abc15a7b4ca41a.pdf.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13–103). New York: American Council on Education/Collier Macmillan.

Mezo, P. G., & Short, M. M. (2012). Construct validity and confirmatory factor analysis of the self-control and self-management scale. *Canadian Journal of Behavioural Science*, *44*(1), 1–8.

Mooi, E., & Sarstedt, M. (2011). *A concise guide to market research: the process, data, and methods using IBM SPSS Statistics*. Springer-Verlag.

Nitko, A. J., & Brookhart, S. M. (2007). *Educational assessment of students*. New Jersey: Pearson Education.

Ramirez, R. P. B., & Ganaden, M. S. (2008). Creative activities and students' higher order thinking skills. *Education Quarterly*, *66*(1), 22–33.

Reynold, C. R., Livingstone, R. B., & Wilson, V. (2009). *Measurement and assessment in education*. Upper Saddle River: Pearson Education Inc.

Rofi'uddin, A. (2000). Model pendidikan berpikir kritis-kreatif untuk siswa sekolah dasar. *Majalah Bahasa Dan Seni*, *1*(28), 72–94.

Shook, C. L., Ketchen, D. J., Hult, G. T. M., & Kacmar, K. M. (2004). An assessment of the use of structural equation modeling in strategic management research. *Strategic Management Journal*, *25*(4), 397–404.

Silverlake, A. C. (1999). *Comprehending test manuals: a guide and workbook*. Los Angeles: Pyrczak Publishing.

Sternberg, R. J. (2006). The nature of creativity. *Creativity Research Journal*, *18*(1), 87–98. https://doi.org/10.1207/s15326934crj1801_10

Subali, B. (2011). Pengukuran kreativitas keterampilan proses sains dalam konteks assessment. *Cakrawala Pendidikan*, *XXX*(1), 130–144.

Suryabrata, S. (2000). *Pengembangan alat ukur psikologis*. Yogyakarta: Andi.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, *56*, 197–208.

Van Dalen, D. B. (1973). *Understanding educational research: an introduction*. McGraw-Hill.

Widowati, A. (2009). Pengembangan critical thinking melalui penerapan model PBL (problem based learning) dalam pembelajaran sains. In *Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA*. Yogyakarta: Fakultas MIPA, Universitas Negeri Yogyakarta.

Wiersma, W. (2000). *Research methods in education: an introduction* (7th ed.). Boston: Allyn & Bacon.