Multitrait-Multimethod Technique for Construct Validity: A Case Study for Instruments of Critical Thinking Ability and Creative Thinking Ability in Programming Course

Ni Made Sri Mertasari[1], Made Candiasa[2]

{candiasa@undiksha.ac.id[1]}

Faculty of Mathematics and Natural Sciences, Universitas Pendidikan Ganesha[1,2]

Abstract. Every psychological measuring instrument has a certain feature or stimulus introduced specifically to represent the trait it intends to measure. A measurement instrument, whether in the form of a test, rating scale, or other devices, almost certainly elicits systematic variance in responses due to the two groups of features. If an irrelevant method variance contributes to the obtained score, then the score is considered invalid. Therefore, it is necessary to test the construct validity of the instrument. This study applied the multitrait-multimethod test to test the validity of the constructs with case studies on creative thinking ability test instruments and critical thinking skills tests in a programming course. Each instrument was developed in two forms, namely description and multiple choice. Initially, the instrument's content validity was tested qualitatively and quantitatively by applying the Lawshe formula. The multitrait-multimethod matrix showed that the convergent validity was met because the reliability values on the diagonals were all high. Furthermore, the multitrait-multimethod matrix also showed that the discriminant validity was also fulfilled because the monotrait-heteromethod cell values showed a high correlation. In contrast, the heterotrait-monomethod cell values showed a low correlation value. The findings above indicated that the creative thinking ability test and critical thinking ability test, which were developed in the form of a description test and a multiple-choice test, have met the construct validity, respectively.

Keywords: multitrait-multimethod, construct validity, creative thinking, critical thinking, programming.

1 Introduction

If a psychological construct is formulated, people often have different thoughts about the formulation of the construct. It is a common phenomenon in social life when an idea is proposed, the surrounding community always has a difference of opinion between supporting new ideas, other ideas, or continuing to use ideas that have been used. One cannot define without implying differences, and verification of these differences is an important part of the validation process. In every psychological measuring instrument, a certain feature or stimulus is introduced

IConVET 2021, November 27, Singaraja, Indonesia

DOI 10.4108/eai.27-11-2021.2315544

specifically to represent the trait it is intended to measure. However, there are not infrequently other features that are characteristic of the method used, and there may even be features used to measure other very different properties. A measurement instrument, whether in the form of a test, rating scale, or other devices, almost certainly elicits systematic variance in responses due to the two groups of features. If an irrelevant method variance contributes to the obtained score, then the score is considered invalid. The source of this invalidity was first formulated by Thorndike and noted as the halo effect (Perera, 2021).

Validity is represented in the congruence between two attempts to measure the same trait through maximally different methods Campbell and Fiske (1959). In this context, it is explained that the reliability of the split-half method is slightly closer to the coefficient of validity than the reliability of the direct re-test because the items are not quite identical. The correlation between different substances may be a measure of reliability, but it is still closer to the area called validity. Campbell and Fiske (1959) then developed a construct validity determination model called the multitrait-multimethod. The multitrait-multimethod analysis is one of the most frequently employed methods to examine the validity of psychological measures (Hintz et al., 2019).

The multitrait-multimethod model was developed based on the study that the validity aspect should emphasize the following specific matters. 1) Validation is generally convergent. Thus, different measurement procedures can confirm it. The independence of the method is the main factor that distinguishes validity, except for content validity. 2) To establish construct validity, discriminant and convergent validity are required. 3) Each measurement is a unity between attributes and methods (trait-method), namely the unity between attributes and measurement procedures not specific to these attributes. The variance between test scores can be caused by responses to measurement methods as well as responses to attributes. 4) To determine discriminant validity and estimate the contribution of method variations, more than one attribute and method must be used in the validation process. In many cases, applying a multitrait-multimethod attribute matrix is very helpful in determining construct validity. The multitrait-multimethod matrix presents all the inter-scores intercorrelation generated when each attribute (content) is measured by each method, as illustrated in Table 1.

Table 1. Reliability of critical thinking ability and creative thinking ability test essay and multiple choice test.

Atribut	Method	Reliability
Critical Thinking Ability (A)	Essay (1)	0.882
	Multiple Choice (2)	0788
Creative Thinking Ability (B)	Essay (1)	0.923
	Multiple Choice (2)	0.825

Magnusson and Backteman (1978) formulated two main requirements for construct validity regarding stability over time, namely: (1) the correlation coefficient of the measurement of the same variable on different occasions must be significantly greater than zero, and (2) the coefficient of stability of the variable must be higher than the correlation between the data of that variable on the first occasion and the data of any variable on the other occasions. In this study, testing of construct validity using multitrait-multimethod was tried to be applied to the

creative thinking ability test instrument and critical thinking ability test in programming learning.

Creative thinking skills and critical thinking skills are in this 21st century. Individuals are required to be able to collect, analyze, and apply information in order to make decisions accurately and quickly. Learning in various fields of study has been directed to the ability to think creatively and critically. However, the instrument for creative thinking is indispensable in producing something new and fulfilling the element of feasibility (Brockman, 1993). Creativity is the main source of all excellence in life (Csikszentmihalyi, 1996). The reason is that all interesting and important things in human life are the result of creativity. Moral values, art, language, science, and technology are the result of creativity, which is obtained through learning activities. Creativity emerges as synergy from several sources, not just one's mind. It is easier to increase creativity by changing environmental conditions than by trying to make people think creatively. Creativity is a component of intellectual life and forms the basis of the development of specific instructional goals in learning ability tests (Guilford, 1967).

Clark (1979) states that creativity is the highest expression of integrated giftedness, synthesizing of all basic human functions. There are four basic human functions, namely taste, senses, ratio, and intuition. The feeling is an effective emotion which is self-actualization. The senses are the creative talent. The ratio is a regular conscious state. Intuition is a conscious condition extracted from the unconscious. The concept includes rational thinking conditions that are measurable and can be developed through various conscious exercises. Everyone is given the freedom to create their own lifestyle.

An individual's lifestyle is shaped by his creativity, and that creativity is responsible for one's life goals, determines methods to fight for life goals, and determines one's interests (Adler in Hjelle & Ziegler, 1992). Creativity also dominates one's perceptions, memories, fantasies, and dreams. Creativity makes humans individuals who are free to determine their destiny. Creativity can exist in humans in various forms, such as writing books, making songs, creating paintings, producing technological products, developing scientific theories, or building organizations that are useful for mankind. Gardner (1993) further states that creativity is the character of a product that at first looks new but is eventually accepted by society. Decisions about creativity can only be determined by people who are experts in the field. Anastasi (1998) adds that the divergent production phase must be followed by a critical assessment to know the level of creativity.

Creativity is a mental activity. Creativity does not occur in the head of a particular individual but rather in the interaction between the human mind and the social-cultural context. So it is more of a systemic phenomenon than an individual phenomenon (Csikszentmihalyi, 1996). This means that creativity is also very much determined by the environment. Setting the right environment for individuals from childhood will play a very important role in shaping the creativity of the individual. The contact between the individual and his environment will provide experiences that can foster the individual's creativity. This statement does not mean that humans must always adapt to their environment because after reaching a certain stage, humans must become architects of their environment. Humans who are able to live well are creative humans, both in ideas, actions, and work. Creative humans easily adapt to their environment because they tend to be constructive in an effort to meet their own needs.

Critical thinking refers to reflective thinking directed at the analysis and evaluation of existing communications, information, and arguments, especially through the use of logic and reason

(Browne & Keeley, 2011). During critical thinking, individuals actively and skillfully conceptualize, apply, analyze, synthesize, and or evaluate information (The Foundation for Critical Thinking, 2021). Information is obtained from observation, experience, or reasoning accompanied by logical evidence. There are two components in critical thinking, namely, the ability to obtain and apply information. For example, technical and vocational education students who are familiar with concepts and interpretations really need critical thinking skills (Reeve, 2016).

Ennis (1985) formulated several individuals who are capable of critical thinking characterized by the ability to define problems, choose criteria for solving problems, formulate settlement plans, make interim decisions, review and make decisions, and monitor implementation. It appears that the indicators of critical thinking ability are very close to the problem-solving steps, namely problem analysis, settlement planning, completion calculations, and completion evaluation (Polya, 1945). Halpern (1998) further states that critical thinking skills are closely related to problem-solving skills, logical reasoning, probability calculations, and decision making.

2 Method

Instruments to measure critical thinking skills and creative thinking skills were tried to be developed. Each instrument was developed in two forms: a description and a multiple-choice test. The description test for creative thinking skills and critical thinking skills consisted of 10 items. On the other hand, the multiple-choice test for both variables consisted of 30 items with five answer choices. The content validity of each test was tested by experts, both qualitatively and quantitatively. Quantitative data from expert test results were analyzed using the Lawshe technique (Lawshe, 1975). The description test is equipped with a rubric that has been tested for consistency with a rater reliability test using the Ebel technique (Ebel, 1972). The multiple-choice test was also tested for the level of difficulty and discriminating power.

To obtain construct validity, all subsequent tests were tested on 30 respondents. The data from the test results per item for each test is used to calculate the reliability of the test using the Cronbach alpha technique. Furthermore, the total score correlation coefficient was calculated using the product-moment correlation technique for each method and trait. The results of the calculation of Cronbach's alpha reliability coefficient (α) and product-moment correlation coefficient (r_{xy}) were then entered into the multitrait-multimethod matrix and interpreted. The interpretation includes the feasibility of the test reliability, convergent validity, and discriminant validity.

Conceptually, the measurement of validity by applying a multitrait-multimethod matrix was carried out using several methods to measure several attributes. The term attribute was used to represent ability or competence when the instrument developed was a test or represent characteristics when the instrument developed was non-test. The multitrait-multimethod validity matrix showed the resulting intercorrelation when each attribute was measured by each method. The classical multitrait-multimethod (MTMM) matrix can be viewed as a two-dimensional cross-classification of traits and methods (Maas, 2009). The intercorrelation between attributes and methods was clearly visible in the multitrait-multimethod matrix, as illustrated in the **Figure 1**.

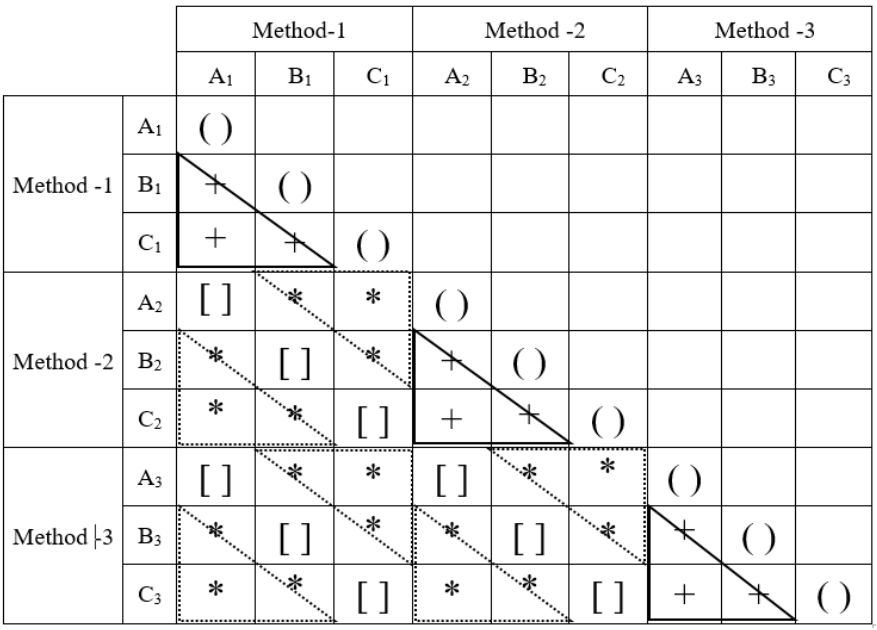

Fig. 1. Multitrait-multimethod matrix illustration.

This illustration involves three different attributes (A, B, and C), each of which is measured by three different methods (Method-1, Method-2, and Method-3), resulting in nine separate variables, which are combinations of three attributes and three methods, namely, attribute A with method-1, attribute A with method-2, attribute A with method-3, attribute B with method-1, and so on until attribute C with method-3. A1 is the attribute for A measured by method-1, A2 is the attribute for A measured by method-2, and so on until C3 is the attribute for C measured by method-3. The confluence of AiAi, BiBi, and CiCi column rows showed the same attribute measured by the same method or called monotrait-monomethod. The confluence of AiAj, BiBj, and CiCj columns showed the same attribute measured by different methods or called monotrait-heteromethod. Furthermore, the confluence of AiBi, AiCi, BiAi, BiCi, CiAi, and CiBi columns showed different attributes measured by the same method or called heterotrait-monomethod. Finally, the confluence of rows and columns of AiBj, AiCj, BiAj, BiCj, CiAj, and CiBj showed different attributes measured by different methods or called heterotrait-heteromethod.

In order to have a more detailed understanding of some of the existing conceptions of the multitrait-multimethod matrix, the explanation in Figure-1 can be seen as follows. The value on the main diagonal, marked by ( ), is the reliability of the instrument to measure each attribute with each method. The value at the upper left end of the diagonal is the reliability of the instrument for measuring attribute A by method-1, and so on until the value at the lower right end of the diagonal is the instrument's reliability for measuring attribute C by method-3. The reliability value on the diagonal can also be referred to as the monotrait-monomethod. Adjacent to the diagonal is a heterotrait-monomethod triangle marked + and bounded by a solid line. The

values in the triangle are the validity of the instrument for measuring different attributes with the same method. The combination of the monotrait-monomethod diagonals and the adjacent heterotrait-monomethod triangles forms a monomethod block. The reason is that all values in the block occur on a monomethod.

The opposite of the monomethod block is the heteromethod which is the instrument's validity value measured by different methods, both for the same or different attributes. Heteromethod block is a combination of diagonal monotrait-heteromethod and heterotrait-heteromethod. The monotrait-heteromethod diagonal contains instrument validity for the same attribute but measured by different methods which are marked in the table [ ]. Meanwhile, the heterotrait-heteromethod triangles are each of two triangles located on each diagonal and bounded by a triangle with a dotted line containing the instrument validity values for different attributes measured by different methods in the table are marked with *. Note that the two heterotrait-heteromethod triangles are not identical.

There are two types of validity in the multitrait-multimethod matrix, namely convergent validity and discriminant validity. Convergent validity appears on the main diagonal which is a monotrait-monomethod correlation. The correlation between the results of measuring the same attribute with the same method should be 1.00 (the correlation of the variable with itself is equal to 1.00). However, in the multitrait-multimethod matrix the coefficients are replaced by the instrument reliability coefficients. If all these coefficients are significantly different from zero and high enough, then the instrument is said to have convergent validity. This information is sufficient to support further validity checks. High-reliability criteria need to get a clear benchmark. Nunnally (1978) limits that for the instruments used in the study, a reliability coefficient of 0.70 is considered high, but for the purposes of instrument standardization, a higher reliability coefficient of 0.90 is required.

Discriminant validity can be seen on several criteria. First, the value on the main diagonal must be higher than the value in the same row or column in the heterotrait-heteromethod triangle. The value of the validity of a variable must be higher than the correlation of the variable with other variables that have no similarities in attributes or measurement methods. These requirements seem self-explanatory and very easy to fulfill, so there's no need to dwell too much on them. However, researchers should remain vigilant because a review of the literature shows that these requirements are often not met, even when the coefficient of validity is quite large. The second criterion is that the correlation of the same attribute measured by different methods is higher than the correlation of different attributes but is measured by the same method. This condition can be monitored in the table by comparing the values in the monotrait-heteromethod validity diagonal in the table marked [ ] with the values in the heterotrait-monomethod triangle in the table marked +. The correlation between scores of instruments measuring the same attribute with different methods should be high because the same attribute measured by any method should produce relatively the same score. On the other hand, the correlation between scores of instruments measuring different attributes using the same method should be low because the instrument has good discriminating power and can measure specific abilities so that different materials measured by the same method have significantly different scores. The third discriminant validity criterion is the pattern of the same attribute relationship shown in all heterotrait triangles in both the monomethod and heteromethod blocks.

Hambleton & Fiske (1959) provide several reasons commonly used to declare discriminant validity invalid, namely low correlation in diagonal validity, too high correlation with other tests intended to measure various things, values in heterotrait-heteromethod triangles as high as values in diagonal validity or even values in the monomethod block, and heterotrait values as high as reliability. Based on the description above, it can be concluded that there are four ideal conditions required of the multitrait-multimethod matrix. The first ideal condition is that the reliability coefficient on the diagonal of the matrix must be high; thus, convergent validity is met. The second ideal condition is that the correlation coefficient between two different methods measuring the same attribute must be high. Furthermore, the third ideal condition is that the correlation between the same method for measuring different attributes must be low. Finally, the fourth ideal condition is that the correlation between different methods for measuring different attributes should be low. The second, third, and fourth ideal conditions guarantee discriminant validity.

3 Results and discussion

The expert test results showed that all tests had met content validity. The test for creative thinking ability and critical thinking ability test, each consisting of 10 items of description test, all met, although there were minor revisions for two test items. Using a quantitative approach with the Lawse technique, the content validity test obtained a content validity ratio (CVR) for all items in the range 0.74-0.83 and a content validity index (CVI) = 0.78. Furthermore, tests of creative thinking skills and critical thinking skills in the form of multiple-choice, each of which consisted of 30 items, it also meets the requirements for content validity qualitatively, although there were still minor revisions to the five test items. Quantitatively with the Lawshe formula, all items have also met content validity, where the content validity ratio (CVR) ranges from 0.75-0.82 and the content validity index (CVI) = 0.78. The results of the content validity test indicate that the developed test was feasible to be tested further.

The reliability coefficient alpha (α) of the instrument for measuring creative thinking skills and critical thinking skills in the form of description tests and multiple-choice tests is as listed in Table 2.

Table 2. Reliability of critical thinking and creative thinking ability test with essay and multiple choice test.

Atribut	Method	Reliability
Critical Thinking	Essay	0.882
	Multiple-Choice	0.788
Critical Thinking	Essay	0.923
	Multiple-Choice	0.825

The correlation coefficient between scores was calculated using the product-moment formula, and the results are as listed in Table 3.

Table 3. Correlation coefficient among total score.

	Critical Thinking with Essay Test	Creative Thinking with Essay Test	Critical Thinking with Multiple-Choice Test	Creative Thinking with Multiple-Choice Test
Critical Thinking with Essay Test	1.000			
Creative Thinking with Essay Test	0.251	1.000		
Critical Thinking with Multiple-Choice Test	0.783	0.378	1.000	
Creative Thinking with Multiple-Choice Test	0.271	0.754	0.239	1.000

The alpha reliability coefficient and the correlation coefficient between scores were then entered into a multitrait-multimethod matrix, as shown in Table 4.

Table 4. Multitrait-multimethod matrix of critical thinking ability test and creative thinking ability test.

Method		Essay		Multiple Choice	
	Atribut	Critical Thinking	Creative Thinking	Critical Thinking	Creative Thinking
Essay	Critical Thinking	(0.882)			
	Creative Thinking	0.251+	(0.788)		
Multiple-Choice	Critical Thinking	[0.783]	0.378*	(0.923)	
	Creative Thinking	0.271*	[0.754]	0.239+	(0.825)

The multitrait-multimethod matrix in Table 4 shows that the convergent validity was met because the reliability values on the diagonals were all high, higher than 0.70. That is, all instruments were feasible to be used to measure data for the measured symptoms. Furthermore, the multitrait-multimethod matrix also showed that the discriminant validity was also fulfilled, proven by the reliability value on the main diagonal being greater than the validity value in the same row and column for the heterotrait-heteromethod triangle (0.271 and 0.378 or those marked *). In addition, the value of monotrait-heteromethod cells showed a high correlation (0.783 and 0.754 or marked [ ]). On the other hand, the heterotrait-monomethod cell values showed a low correlation value (0.251 and 0.239 or the + sign). Thus, all discriminant validity criteria were met. The results of the findings indicated that the creative thinking and critical

thinking ability test, which were developed in the form of a description and a multiple-choice test, had met the construct validity, respectively.

4 Conclusion

Test instruments to measure creative thinking skills and critical thinking skills have been developed respectively in the form of essay tests and multiple-choice tests. All items of the instrument have met the requirements of content validity. In addition, all instruments also meet the reliability requirements. Another important finding was that the instrument had met construct validity. The applied multitrait-multimethod technique showed that all instruments met convergent validity and discriminant validity. Thus, the test instruments for creative thinking skills and critical thinking skills tests were feasible to be applied. The findings above showed that the ability to think creatively and the ability to think critically have a close relationship. Both support each other as individuals seek to solve problems, express themselves, innovate, improvise, and socialize. Therefore, when the measurement was carried out with different instruments, the construct validity was still fulfilled.

Creativity is the highest expression of integrated talent, synthesizing all basic human functions, namely taste, senses, ratio, and intuition. Creativity also dominates one's perceptions, memories, fantasies, and dreams. Creativity has the potential to exist in humans in various forms, thus making humans individuals who are free to determine their destiny. Critical thinking refers to reflective thinking directed at the analysis and evaluation of existing communications, information, and arguments, primarily through the use of logic and reason. During critical thinking, individuals actively and skillfully conceptualize, apply, analyze, synthesize, and or evaluate information. Individuals who are able to think critically are characterized by the ability to define problems, choose criteria for solving problems, formulate plans, make decisions, review, and monitor implementation.

Efforts still need to be made to test the construct validity of all test instruments developed using confirmatory factor analysis (CFA). If the CFA test also shows that the construct validity has been met, then the findings are more convincing for users to apply the instrument. If these efforts have occurred, then the next effort needed is to develop other instruments. For further, a question bank is expected to test creative and critical thinking skills. The question bank is realized in digital form so that each time it is used, it can present a different device. Thus the test result bias due to history or experience can be avoided and the test results obtained are more authentic.

References

[1] Ananstasi, Anne & Susana Urbina. (1988). *Psychological Testing.* New Jersey: Prentice Hall

[2] Brockman, John. (1993). *Creativity*, New York: Simon & Schuster

[3] Campbell, D. T. & Fiske, D. W., "Convergent and discriminant validation by the multitrait-multimethod matrix" *Psychological Bulletin*, 1959, 56, 81-105.

[4] Csikszentmihalyi, Mihaly. (1996). *Creativity.* New York: HarperCollins Publishers

[5] Ebel, Robert L. (1972). *Essential of Educational Measurement,* Englewood Cliff, NJ: Prentice Hall, Inc.

[6] Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. E*ducational Leadership, 43*(2), 44–48.
[7] Gardner, Howard. (1993). *Multiple Intelligences.* (New York: Basic Books
[8] Guilford, J.P. (1967). Creativity: Yesterday, Today and Tomorrow. *Journal of Creative Behavior*. Volume 1, Issue 1, Winter 1967, PP. 3-14
[9] Halpern, D. F. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449–455.
[10] Hintz, Fred A., Christian Geiser, G. Leonard Burn, and Mateu Severa. (2019). Examining Quadratic Relationships Between Traits and Methods in Two Multitrait-Multimethod Models. Front. Psychol., 14 March 2019 | https://doi.org/10.3389/fpsyg.2019.00353
[11] Hjelle, Larry A. & Daniel J. Ziegler. (1992). *Personality Theories.* Singapore: McGraw-Hill
[12] Lawshe, C. H. (1975). "A Quantitative Approach to Content Validity", Personnel Psychology, 1975, 28, 563-575.
[13] Maas, Cora J. M., Gerty J. L. M. Lensvelt-Mulders, and Joop J. Hox. (2009). A Multilevel Multitrait-Multimethod Analysis. Methodology 2009; Vol. 5(3):72–77. DOI: 10.1027/1614-2241.5.3.72
[14] Magnusson, D. & G. Backteman. (1978). "Longitudinal Stability of Person Characteristics: Intelligence and Creativity", Applied Psychological Measurement, Vol. 2, No. 4 Fall 1978, pp. 481-490.
[15] Nunnally, J. C. (1978). *Psychometric Theory,* New York: McGraw-Hill Book Company
[16] Perera, Ayesh. (2021). Why the Halo Effect Affects How We Perceive Others. *Simply Psychology*, March 22, 2021. From https://www.simplypsychology.org/halo-effect.html
[17] Polya, G. (1945). *How to Solve It.* Princeton: Princeton University Press.
[18] Reeve, E. M. (2016). 21st-century skills needed by students in technical and vocational education and training (TVET). *Asian International Journal of Social Sciences, 16*(4), 65 – 82. doi: 10.29139/aijss.20160404