ARTIKEL PENELITIAN

Development and Validity of Fluid Intelligence Test Based on Cattle-Horn-Carrol Theory: A Pilot Project

FITRI ANDRIANI[1*], CHOLICHUL HADI[1] & PRAMESTI PRADNA PARAMITA[1]

[1] Fakultas Psikologi, Universitas Airlangga

ABSTRACT

This study aimed to examine the validity of the Fluid Intelligence Test, constructed based on the Cattel-Horn-Carroll theory. There were two sources of validity used in this study, which were evidence based on the internal structure and evidence based on relation with other variables. Sixty-four items have been composed and tested to 242 people. The data was analyzed using confirmatory factor analysis technique and correlations technique to examine test validity. The result of this study showed that the prepared model worked quite well in describing the narrow abilities of fluid intelligence, as showed by the receipt of fit indexes accuracy of the model, such as Chi-Square value .42 (p = .518), Goodness Fit Index (GFI) = 1, and Rooted Mean Square Error (RMSEA) = .00. Similar result was also showed by its correlation with other variables, which are .717 (TIKI) and .606 (CFIT). This suggested that Fluid Intelligence Test has good validity.

Key words: *fluid intelligence, test development, test validity.*

INTRODUCTION

Every individual possesses his/her own uniqueness, or the so-called individual differences. Individual differences consist of physical and psychic differences. Physical differences refer to body shape, stature, skin color, hair color, blood type, or other physical attributes. While psychic differences refer to the psychological aspect in one individual, both cognitive and non-cognitive; such as personality, motivation, adaptability, interactive ability, numeracy, memory, understanding ability, analytical ability, problem solving ability, and intelligence. Intelligence indicates one's quality and it is essential to find out one's quality and uniqueness. There are several reasons why testing one's intelligence has become essential. For instance, in decision making related to one individual or a group of individuals, the information on each individual's intelligence level is needed. Obviously, a good decision is derived from

an accurate data and information. Meanwhile, psychological ability assessment is one of the ways to gather information using a range of psychological measurement. In making decision, one's intelligence is an important determinant. For instance, in order to determine whether one should be hired or not, a psychologist/company considers his/her intelligence. Moreover, in clinical field, intelligence becomes one of the consideration to determine some diagnoses. Similarly, in educational field, a decision-making process in which intelligence becomes the main consideration is commonly done. For example, to decide whether one's eligible for education scholarship, he/she is required to possess a high intelligence level. Intelligence test is also employed to test whether a student is eligible to enroll in acceleration program or not. Additionally, some selection programs which are commonly conducted by various educational level also required intelligence test; particularly due to the limited number of seats compared to the overwhelming applicants' number.

In order to find out individuals' psychological attributes, like intelligence, we can employ several tests. As a measurement tool, intelligence tests determine the accuracy of the information. Therefore, a representative test is essential to find out what aspects indicate individual differences. Furthermore, related to the need to measure one's intelligence, we need a tool to measure standard intelligence which fulfil the required psychometric property. It shall be able to be theoretically accountable, prioritize ethical codes, and fulfil the psychometrical requirements (validity and reliability). If the measurement tool is unable to fulfil the aforementioned standards, the results shall be doubtful as well. Moreover, if the doubtful results are utilized to make a decision, the decision shall be wrong, misleading, and detrimental for various parties, both users and the individual concerned.

In order to measure intelligence, we need a representative tool; however, in reality, such tool is somewhat limited. The limitation is derived from the restricted variation of intelligence tests. Another related problem in measuring intelligence is its wide scope of use. It is utilized for various objectives by the psychologists so that it becomes highly popular; however, as it is too frequently used, it demands new variety of tools as well. In Indonesia, the measurement tool is highly limited; thus, one tool is used repeatedly and it affects its validity. In conclusion, the problems related to psychological assessment activities are listed as follows; 1) the more frequent a tool is used, the more decreasing its validity shall be; 2) limited test variety in Indonesia; 3) urgent need to renew or develop a new tool as an alternative for the existing one. Based on such problems, this study aims to solve it by developing a test to measure cognitive ability, called fluid intelligence. It also aims to conduct a validity test for the newly composed test.

Definition of Intelligence

There are various definitions of intelligence from many scholars according to their own perspectives; such as Binet, Spearman, Thorndike Thurstone, Weschler, Gardner, and Sternberg. There is no definite truth for one theory which sets it apart from other, because intelligence is intangible and can only be viewed through some indicators and actualized characteristics of individuals. Binet-Simon defined intelligence as a three-components contract; a) ability to direct one's mind or action; b) ability to change the direction if an action is taken; c) ability to conduct self-critics. The second definition comes from Spearman and Jones. They stated that intelligence is an ability to do something in which some general factors (g) and specific factors (s) are included. Third, Thorndike, et al. (1955) stated that intelligence is an ability to generate a good response based on some facts or truth (Wilson, 1974). Meanwhile, Thurstone offered a different view from Thorndike. He said that intelligence is derived from six abilities: verbal, numeric, spatial, word fluency, memory, and reasoning. Fifth, Gardner offered his famous multiple intelligence theory. He proposed that a human possesses different kinds of intelligence, such as linguistic, numeric, spatial, musical, bodily kinesthetic, interpersonal, and intrapersonal. Sixth,

Weschler (1958) affirmed that intelligence is a set of one's ability to act according to certain objectives, think rationally, and face his/her surroundings effectively. Finally, by integrating the previous theories, Sternberg (1982) conducted a research to find deeper definitions of intelligence. He then concluded that intelligence is comprised of three main abilities: a) ability to solve practical problems using logical way of thinking; b) verbal ability which is characterized by eloquency; and c) social competency which included the ability to accept others the way they are. Meanwhile, the definitions stated by Binet-Simon, Spearman, Thorndike, Gardner, Wechler, dan Sternberg are cited in Azwar's (1999) study.

Meanwhile, Cattell, Horn, and Carrol presented a theory of intelligence which is frequently shortened as CHC Theory. Cattell (1941, 1971) initiated the formulation of the theory, then it was revised by John Horn (1968) and Carrol (1993). Carrol (1993) reviewed and analyzed a thousand of intelligence test's database and resulted in three startum models of human intelligence and cognitive abilities. This model is considered as the best in describing human's cognitive structure due to its powerful empirical foundation. Carrol's theory was then reviewed by McGrew and Flanagan. Both suggested the integration of Cattel-Horn's and Carrol's theories; thus, the current CHC theory was created. CHC is a taxonomy model which integrates various factor analysis research on intelligence. Those research dated back to 100 years ago. CHC has the broadest implication on intelligence measurement (McGrew, 1997; Gregory, 2011). According to CHC theory, intelligence is comprised of pervasive, broad, and narrow ability which are structured hierarchically. Table 1 presents the theory scheme based on CHC. On the other hand, according to Carrol (1993), McGrew (1997) and www.iapsych.com (as cited in Gregory, 2011), fluid intelligence has the following meaning: Fluid intelligence/reasoning (Gf) comprises of high level reasoning. It is used to perform new tasks which cannot be done automatically. Mental operation in fluid intelligence includes the ability to draw a conclusion, shape a concept, compose a hypothesis, understand implications, as well as reason inductively and deductively. The narrow ability of fluid intelligence is explained at Table 2.

Table 1. Theory scheme based on CHC (Three Stratum Theory)

Stratum III	Stratum II	Stratum I
General Intelligence (g)	Fluid inteligence	5 narrow ability
	Christalized intlligence	10 narrow ability
	Domain-specific knowledge	7 narrow ability
	Visual spatial abilities	11 narrow ability
	Auditory processing	13 narrow ability
	Broad retrieval / memory	13 narrow ability
	Cognitive processing speed	7 narrow ability
	Decision / reaction time or speed	5 narrow ability

Source : Mc Grew (2003), Gregory (2011)

Table 2. Narrow ability dari fluid intelligence (Mc Grew, 2005)

Ability	Description
Fluid intelligence / reasoning (Gf)	The use of deliberate and controlled mental operations to solve novel, "on-the spot problems (i.e., tasks that cannot be performed automatically). Mental operations often include drawing inferences, concept formation, classification, generating and testing hypotheses, identifying relations, comprehending implications, problem solving, extrapolating, and transforming information. Inductive reasoning (inference of a generalized

Ability	Description
	conclusion from particular instances) and deductive reasoning (the deriving of a conclusion by reasoning; specifically, inference in which the conclusion about particulars follows necessarily from general or universal premises) are generally considered the hallmark indicators of Gf. Gf has been linked to *cognitive complexity*, which can be defined as a greater use of a wide and diverse array of elementary cognitive processes during performance.
General sequential (deductive) reasoning (RG)	Ability to start with stated assertions (rules, premises, conditions orientation) and to engage in one or more steps leading to a solution to a problem. The processes are deductive as evidenced in the ability to reason and draw conclusions from given general conditions or premises to the specific. Often known as *hypothetico-deductive reasoning.*
Induction (I)	Ability to discover the underlying characteristic (e.g., rule, concept, principle, process, trend, class membership) that underlies a specific problem or a set of observations, or to apply a previously learned rule to the problem. Reasoning from specific cases or observations to general rules or broad generalizations. Often requires the ability to combine separate pieces of information in the formation of inferences, rules, hypotheses, or conclusions.
Quantitative reasoning (RQ)	Ability to inductively (I) and/or deductively (RG) reason with concepts involving mathematical relations and properties.

Source : McGrew (1997)

Intelligence Test

As suggested by its name, intelligence test is a psychological test which aims to measure one's intelligence. In psychology, intelligence test plays an essential role, as intelligence itself is an vital attribute within oneself. There are various intelligence test which is created by many scholars, along with their underlying theories; such as WAIS, Binet, CFIT, WAIS, WISC, WPPSI, KAIT, K-ABC, STANFORD-BINET, PASS, NNAT, IST, RAVEN, DAS, WJ-COG, DAS-DNCAS, UNIT, RIAS, etc. In Indonesia, the intelligence test which has been adapted to Indonesian, rather popular, and frequently used are WAIS, Binet, CFIT, WAIS, WISC, WPPSI, and IST. In addition, there is an intelligence test which is made by Indonesian, called TIKI (Indonesian Collective Intelligence Test). As previously mentioned, compare to the development of intelligence test above, the intelligence tests in Indonesia are highly limited. The limitation (either in nationally-composed test or the adapted ones) has become the reason to develop a new intelligence test as an alternative.

Classifying Intelligence Test

Basically, intelligence test can be divided into two: based on the test-takers' number and the item format. Based on the test-takers' number, it can be classified into individual test and classical test. Individual test refers to one-on-one test for one individual; for example, CFIT, IST, etc. While the classical test can be taken by a group of people; for example, BINET, WAIS, etc.

Furthermore, based on the completion method, intelligence test is categorized into verbal and non-verbal test. Verbal test instructs the test-takers to comprehend verbal concepts and answer the items

verbally. Otherwise, non-verbal test does not require the takers to understand verbal concepts and is conducted by either performance or paper and pencil method.

METHOD

There are 70 items, which were tested for trials three times. The first trial was a preliminary study for 34 subjects. From the preliminary study, a selection and item revisions are conducted. Next, an analysis is conducted for the second trial using ITEMAN in order to obtain item statistic and scale statistic. Further, selection and item revision were performed based on the previous analysis; 64 items are obtained and tested on 242 subjects. The last trial involved larger group contained 699 subjects using 52 remaining items. Validity test was executed by using two methods: criteria-based method and confirmatory factor analysis method. The first method was performed by correlating TIKI and CFIT total score by using product moment correlation technique.

RESULT AND DISCUSSION

Based on the data processing results using ITEMAN in the first trial, scale statistics summary is described in Table 3, while the second trial results are depicted in Table 4 and the last trial result is presented in Table 5. Meanwhile, the validity test results for the correlation approach (the correlation between fluid intelligence and TIKI test) is included in Table 6, while correlation with CFIT test is described in Table 7. Furthermore, through the confirmatory factor analysis, the validity test is presented in Table 8.

Table 3. First Trial Fluid Intelligence Statistics Summary (N=34, 70 items)

Statistic scale	Scale
N of Items	70
N of Examinees	34
Mean	46.382
Variance	64.236
Std. Dev.	8.015
Skew	-.313
Kurtosis	-.622
Minimum	31.000
Maximum	63.000
Median	48.000
Alpha	.819
SEM	3.413
Mean P	.663
Mean Item-Tot.	.265
Mean Biserial	.359

The selection for good item (in terms of fulfilling psychometrical property for this activity) can be based on discrimination power and level of difficulty. From the data analysis on 70 items which are tested to 34 subjects, the temporary results of 47 items with discrimination power above .3 are obtained. The whole statistical results can be viewed based on the Cronbach Alpha, SEM, *p* mean, total mean item, and bi-serial mean's results. Cronbach Alpha's result (.819) is obtained from a reliable measurement tool. SEM presents moderate error margin which is obtained from the measurement of standard deviation and reliability; likewise, the data analysis' result for trial test show similar results. Meanwhile, the *p* mean's result indicates average level of difficulty (.663); proving that the items in this measurement test

has medium difficulty level. Furthermore, the total mean item recorded as .259 and bi-serial mean is recorded .359; implying that the items are capable of differentiating the subjects from high level ability to low level ones.

Table 4. Second Trial Fluid Intelligence Statistics Summary (N=242, 64 items)

Statistic scale	Scale
N of Items	64
N of Examinees	243
Mean	46.107
Variance	70.079
Std. Dev.	8.371
Skew	-1.387
Kurtosis	3.488
Minimum	20.000
Maximum	59.000
Median	48.000
Alpha	.858
SEM	3.149
Mean P	.720
Mean Item-Tot.	.321
Mean-Biserial	.46570

Meanwhile, based on the results of the obtained 47 items, the revision and addition of items are conducted during the second trial test with larger number of subjects; in which 64 final items for the second trial are acquired. Furthermore, the statistical summary based on the data processing using ITEMAN is presented in Table 3. As described in the discussion section for the first trial in this research, the selection for good items (which fulfil psychometrical property) is constructed by discriminating power and level of difficulty. There are 52 items which possessed discrimination power above .3. The statistical measurement (as depicted in Table 4) can be obtained from Cronbach Alpha, SEM, p mean, total item mean, and bi-serial mean. Cronbach Alpha's result (.858) shows the reliable measurement result. While SEM presents a moderate error based on the measurement of standard deviation and reliability; this result is also reflected in the trial test's results. Furthermore, p mean indicates a medium difficulty level (.720); while total mean item (.321) and bi-serial mean (.466) denote the ability to discriminate the high ability level and low ones.

Table 5. Third Trial Fluid Intelligence Statistics Summary (N=699, 52 items)

Statistic scale	Scale
N of Items	52
N of Examinees	699
Mean	35.682
Variance	60. 088
Std. Dev.	7.752
Skew	-0.635
Kurtosis	-0.183
Minimum	10.000
Maximum	50.000
Median	37.000
Alpha	0.857
SEM	2.929

Statistic scale	Scale
Mean P	0.686
Mean Item-Tot.	0.344
Mean Biserial	0.476

Then, the third field testing on 52 items which are tested to 699 subjects (Table 5). The whole statistical results can be viewed based on the Cronbach Alpha, SEM, *p* mean, total mean item, and bi-serial mean's results. Cronbach Alpha's result (.857) is obtained from a reliable measurement tool. SEM presents moderate error margin which is obtained from the measurement of standard deviation and reliability; likewise, the data analysis' result for trial test show similar results.

Furthermore, the total mean item recorded as .344 and bi-serial mean is recorded as .476; implying that the items are capable of differentiating the subjects from high level ability to low level ones. The result was better than the first trial and second trial. While SEM presents a moderate error based on the measurement of standard deviation and reliability; this result is also reflected in the trial test's results. Furthermore, *p* mean indicates a medium difficulty level (.686); while total mean item (.344) and bi-serial mean (.476) denote the ability to discriminate the high ability level and low ones.

Table 6. Correlation Between Fluid Intelligence Test and TIKI test (N=242)

		TIKI	FLUID
TIKI	Pearson Correlation	1	.717**
	Sig. (2-tailed)		.000
	N	242	242
FLUID	Pearson Correlation	.717**	1
	Sig. (2-tailed)	.000	
	N	242	242

**. Correlation is significant at the 0.01 level (2-tailed).

Table 7. Correlation between Fluid Intelligence Test and CFIT test (N=699)

		TIKI	FLUID
CFIT	Pearson Correlation	1	.606**
	Sig. (2-tailed)		.000
	N	699	699
FLUID	Pearson Correlation	.606**	1
	Sig. (2-tailed)	.000	
	N	699	699

**. Correlation is significant at the 0.01 level (2-tailed).

The criteria-based validity was obtained by correlating Fluid Intelligence test with TIKI and CFIT. In accordance with Table 6, there are 242 subjects and each of them possess two kinds of score: TIKI and fluid intelligence. The correlation result is shown at .717 (p<.000); and in accordance with Table 7, there are 699 subjects and each of them possess two kinds of score: CFIT and Fluid Intelligence. The correlation result is shown at .699 (p<.000). These indicating significant correlation between TIKI and Fluid Intelligence's test score, and also between CFIT and fluid intelligence's test score.

Table 8. Model Fit Indexes with SEM

Fit Indexes	The range be expected	Model's Index	Description
Chi Square	Small	.42	Good
p-value	>.05	.518	Good
GFI	>.90	1	Good
RMSEA	<.08	.00	Good
CFI	>.94	1	Good

Data presented in Table 8 proves that Fluid Intelligence Test is appropriate in measuring intelligence. Table 8 presents the results of internal validity based on confirmatory factor analysis (CFA). CFA's results prove that the model presented in this study represents the fluid intelligence well (Chi-Square .42; p=.518; Goodness Fit Index [GFI] = 1, Rooted Mean Square Error [RMSEA] =.00). Likewise, the correlation with other variables (TIKI test) is recorded at .717, further proving that Fluid Intelligence Test possess a good validity.

CONCLUSION

This study concludes that it has successfully creates a measurement tool for intelligence test. Based on the trial test's analysis results, there are 52 items which fulfil psychometrical property and validated based on the correlation with other variables and confirmatory factor analysis. The validation criteria's results which are obtained by correlating Fluid Intelligence Test with TIKI test offer a significant correlation; thus, it implies that fluid intelligence is capable of measuring intelligence. Similarly, a good result for goodness of fit index during the confirmatory factor analysis confirms that Fluid Intelligence Test is structurally capable of measuring intelligence constructs.

Finally, this research suggests a prudent usage of fluid intelligence, as this measurement tool is only tested as a pilot project. A further trial with more heterogeneous subjects is needed. It is also possible to add more items which are followed by standardization and norm development.

REFERENCES

Azwar, S. (1999). Pengantar Psikologi Inteligensi. Yogyakarta: Pustaka Pelajar Offset.

Carrol, J. B. (1993). Human Cognitive Abilities: A Survey of factor Analysis Studies. Australia: Cambridge University Press.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, *38*(592), 10.

Cattell, R. B. (1971). *Abilities: their structure, growth and action.* Oxford: Houghton Mifflin.

Gregory, R. J. (2013). Tes Psikologi: Sejarah, Prinsip dan Aplikasi. Edisi 6. Terj. Jakarta: Penerbit Erlangga.

Horn, J. L. (1968). Organization of Abilities and The Development of Intelligence. *Psychological Review, 75*(3), 242-259.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. Dalam D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Ed.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York: Guilford.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. Dalam D. P. Flanagan & P. L. Harrison (Ed.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed.). New York: Guilford.

Thorndike, R. L. & Hagen, E. (1955). Measurement and Evaluation in Psychology and Education. New York: John Willey and Sons.

Wechsler, D. (1958). The Measurement and Appraisal of Adult Intelligence

Wilson, R. S. (1974). Twins: Mental development in the preschool years. *Developmental Psychology*, *10*(4), 580.