

The Development of Visualization Test Based on Cattel-Horn-Carroll (CHC) Theory of Intelligence for Senior High School Student in Indonesia

Andhika Prakoso^a and Nurul Arbiyah^{b*}

^a*Faculty of Psychology, Universitas Indonesia, Depok, Indonesia;* ^b*Psychology Research Method Department, Faculty of Psychology, Universitas Indonesia, Depok, Indonesia*

*Corresponding Author:

Nurul Arbiyah

Psychology Research Method Department
Faculty of Psychology, Universitas Indonesia
Jl. Lkr. Kampus Raya, Depok, Jawa Barat
Indonesia, 16424

Tel.: +62 217270004

Email address: nurul.arbiyah31@ui.ac.id

The Development of Visualization Test Based on Cattel-Horn-Carroll (CHC) Theory of Intelligence for Senior High School Student in Indonesia

Visual-spatial ability measurement for the recommendation for 10th grade high school students serves as the basis for constructing a visual-spatial test. Cattell-Horn-Carroll theory, which is currently the most comprehensive theory of intelligence, is used as the structural framework of this test. This research attempts to construct a visual-spatial test that contains items that discriminate individual ability, has distractors that perform well and are spread evenly among false options, and can be meaningfully interpreted. This research also attempts to construct a reliable and valid measure of visual-spatial ability. This test was administered to 97 10th grade high school students from three high schools in Jakarta ($M_{age}=16,01$; $SD = 4,691$). Results show that visual-spatial measurement is internally consistent and is correlated significantly with subtest 2 of TIKI M (Combined Parts) and subtest 7 of TIKI M (Forming Shapes), thereby indicating the validity of this test. Initially, this test consisted of 35 items. However, several items that did not meet the criteria were eliminated. The final version of this test has 20 items with varying degrees of difficulty, can discriminate an individual's ability, and has distractors that perform optimally. The visual-spatial test score was interpreted using within-group norms with a scaled score. This test is proven to be a suitable visual-spatial measurement tool that can be used for recommendation for 10th grade high school students.

Keyword: Cattel-Horn-Carroll Intelligence Theory; High school Major; Test Construction; Visualization

Introduction

The challenges of globalization and competition between countries are now becoming an issue of concern for the government because Indonesian students currently do not have the qualities to undertake such challenges yet (Kemdikbud, 2014). Therefore, the government established the 2013 curriculum (K13), which is designed to develop the competency and ability that students need to respond to such challenges. This curriculum includes the specialization of senior high school students to develop their potential in relevant learning programs. Students enter senior high school major programs based on their interest, talent, and/or academic ability (Kemdikbud, 2013). However, many students are still confused as to which program matches their interest, talent, and/or academic ability. Thus, students end up choosing programs based on their parents' recommendations or what their friends choose and not on their interest (Rozi & Irawan, 2015; Dewi & Nuryono, 2014). As a result, students tend to lose interest in learning, which results in poor academic achievement (Rozi & Irawan, 2015). Therefore, each student must join a major program based on their interest, talent, and/or academic ability.

One of the major factors that determine students' scholastic interests is academic ability. Academic ability is a series of basic and complex cognitive abilities, such as the ability to read, perform math, or practice critical thinking, which act as a determinant of learning success (DiPerna, 2006; DiPerna, 2008). Integrating cognitive ability in major programs requires students to have or develop particular cognitive abilities throughout the course of the program. If students possess those cognitive abilities, then they have the basic tools that can help them learn and comprehend school subjects.

DiPerna (2008) explained that the level of a person's cognitive abilities needs to be known to serve as a basis for further educational development. Unfortunately, the current major program selection system in Indonesia is unable to fulfill this need. At present, the major program selection system is based on three predictors: the national exam, junior high school reports, and recommendations of counseling teachers. However, these three predictors are still lacking. A literature review of the national exam and student's reports showed that other measured factors exist aside from cognitive ability (Sirin, 2005; Fan & Chen, 2001; Poropat, 2009; Multon, Brown, & Lent, 1991; Sanders, Wright, & Horn, 1997; Seipp, 1991). The recommendations of counseling teachers cannot address the limitations of the two previous predictors because no clear and standardized procedures exist with regard to how recommendations are given. Thus, numerous ways of giving recommendations exist, which are used to help school officers decide the best programs for each student. Current selection systems measure not only cognitive ability, which is necessary for further student educational development, but also other factors. Thus, this system needs to be improved with the addition of accurate, consistent, and standardized data to measure students' cognitive abilities.

One way to tap students' cognitive ability is to conduct an intelligence test. An intelligence test assesses an individual's intelligence as a whole (Benson, 2003). It can provide an overview of potential cognitive capacity and assist in placement and selection by detecting a person's ability (Kaufman & Lichtenberger, 2006; Cohen & Swerdlik, 2009). Through intelligence measurement, a school obtains an overview of a student's potential and cognitive capacity, thereby increasing the accuracy of the specialization assessment. Schneider and McGrew (2012) explained that Cattell–Horn–Carrol (CHC) intelligence theory is currently the most comprehensive model in describing human cognitive structures. The taxonomy theory of CHC consists of three strata, namely, general intelligence or G factor, broad abilities, and narrow abilities. The strata constitute a taxonomy in the form of levels, with G factor at the highest level. In relation to placement requirements, conducting intelligence measurements based on the narrow abilities strata is recommended (McGrew & Wendling, 2010). However, a question arises: What narrow abilities strata on CHC theory should be measured as a predictor for major program selection?

The essential cognitive ability in the CHCs intelligence taxonomy for senior high school specialization students is visual–spatial ability (Vz). Visual–spatial ability is the core of the broad ability visual processing (Gv) or spatial abilities. Research has shown that visual–spatial ability is a good predictor for individuals who enter the science, technology, engineering, and mathematics (STEM) field (Wai, Lubianski, & Benbow, 2009; Johnson, Barron, Rose, & Carreta, 2017). Wai et al. (2009) stated that visual–spatial ability is an important psychological attribute for students to enter an advanced STEM program. A longitudinal study showed that individuals with good visual–spatial ability successfully achieved PhDs in STEM (Wai et al., 2009). The STEM field requires projection and spatial understanding (Uttal & Cohen, 2012), and individuals with good visual–spatial ability can efficiently perform tasks that require advanced spatial understanding (Uttal & Cohen, 2012).

Recent research explained that visual–spatial ability is the most significant factor in predicting success in the STEM field when individuals have not received domain-related knowledge (Johnson et al, 2017). Visual–spatial ability is the core skill needed by an individual to learn and comprehend knowledge related to the STEM field. STEM is one of the early education paths for students with good visual–spatial ability who aim to enter the STEM field. Senior high school

students with good spatial skills are predicted to be able to succeed in the STEM field, thereby achieving the goal of the 2013 curriculum to produce qualified individuals.

In Indonesia, few intelligence tests have been developed based on CHC theory. Currently, one intelligence test based on CHC theory is being developed by Yayasan Dharma Bermakna, 2015). However, their constructed intelligence test is designed for younger children and needs to be administered individually. Thus, this test is not relevant to the current study, is less efficient, and requires a considerable amount of time to administer. Hence, a new spatial ability testing tool based on CHC theory and can be administered to several students at once needs to be developed. The test to be constructed is expected to provide an overview of the spatial ability of senior high school students who will be recommended to enter the science specialization. On the basis of these tests, the school will have another baseline for accurate decision making based on the student's abilities, especially in science specialization.

Thus, on the basis of CHC theory, this research aims to construct a visual–spatial ability test that is consistent, valid, has gradual difficulty items according to the purpose of the test, is able to distinguish the visual–spatial ability of the testers, has an optimal distractor answer option, and is able to give meanings to individual scores.

CHC Theory of Intelligence. Schneider and McGrew (2012) explained that CHC theory of intelligence is the result of the integration between the leading intelligence theories of Cattel–Horn's Gf–Gc theory and Carroll's three stratum theory. CHC theory consists of three levels that make up human cognitive structures. The third level is the highest strata, which represents the concept of the G factor. The second level covers broad abilities, which are the characteristics or abilities of the individual that influence and are influenced by the daily behavior of the related domain (Carroll, 1993). The first level is the lowest structure, which covers narrow abilities, referring to a specialization ability that reflects the effects of experience and the learning process or the adoption of specific strategies (Carroll, 1993). Currently, the broad abilities level consists of 16 domain capabilities, and the narrow abilities structure consists of approximately 80 capabilities (Schneider & McGrew, 2012).

Visual–spatial ability. Visual–spatial ability is the ability to form the broad ability visual processing (Gv). Visual processing (Gv) is the ability to use stimulated mental imagery (generally beginning with visual perception) to solve problems with a given image stimulus (Schneider & McGrew, 2012). The domain of this ability consists of 11 narrow abilities compilers, where visual–spatial ability becomes the core ability of the compiler (Schneider & McGrew, 2012).

Visual–spatial ability is the ability to perceive complex patterns and mentally simulate how an object might look when transformed (e.g., rotated, resized, and partially displayed) (Schneider & McGrew, 2012). Mental stimulation is the process of forming representations of shapes, images, or positions and manipulating the visual form mentally (Carroll, 1993). Referring to the explanation, the core of this ability is how one person perceives the visual stimulus and uses mental simulations to see how the visual pattern will look in certain conditions. The visual–spatial ability lies in the process of mental simulation performed by the individual to process given stimulus.

Schrink, McGrew, and Mather (2014) explained that visual–spatial ability has two dimensions, namely, space relation and block rotation. Space relation is a representation of visualization ability and involves identifying two-dimensional fractions into a complete form. Block rotation is a representation of visual–spatial ability by performing mental simulations on the building of three-dimensional space such as cubes, beams, and prisms. This dimension will be used as a reference in the test construction.

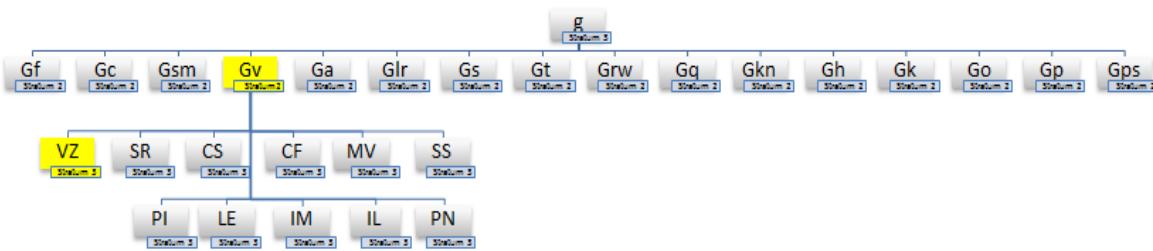


Figure 1. CHC taxonomy theory to be constructed.

Methods

Participant

Ninety-seven 10th grade students from Jakarta (33 male respondents and 64 female respondents) participated in this study. Their mean age was 16.01 years old. The respondents were distributed between two programs: 35 respondents from the science program and 62 respondents from the social science program.

Measures

A visual–spatial test includes the space relation and block rotation (Schrink et al, 2014). Test items also consider the level of difficulty, which is based on item complexity (Ramful, Lowrie, & Logan, 2016). In the space relation dimension, the number of fractions determine the level of difficulty, that is, more fractions correspond to increased difficulty in solving the problem. The difficulty in the block rotation dimension is determined based on the rotation degree of items. A large required rotation degree corresponds to increased difficulty in solving the problem. Items with a rotation degree below 90° are easier to solve than items with a rotation degree >90°.

The space relation dimension consists of 16 items, and the block rotation dimension consists of 19 items. This test has a total of 35 pooled items and 20 desired items. Fifteen items with less desired characteristics are eliminated. This decision considers practicality and test takers' weariness.

Scores obtained by test takers are a sum of every correctly answered item or cumulative model (Cohen & Swerdlik, 2009). Every correctly answered item receives one point, whereas wrong items receive zero points. High scores indicate high visualization ability.

Procedure

The construction procedures are divided into several steps: conceptualization, construction, expert judgment and test legibility, test trial, and test revision (Cohen and Swerdlik, 2009). Conceptualization was conducted to match the needs and purpose of the test and CHC intelligence theory. The conceptualization stage produces the basic framework and the constructed visual-spatial ability dimension.

Test construction is the process of developing the test items and involves determining the form of items (multiple choice), operationalizing construct visualization (Vz) into operational forms that are ready to be used in the item making process, and writing instructions to be read by the respondents.

The next procedure is expert judgment and legibility test. Expert judgment is performed to determine whether the items represent the spatial ability's behavior domain. Experts were chosen for their expertise in intelligence test construction. The process involves examining whether or not every item on the test corresponded to the indicator of sample behavior. If the experts determined that an item does not correspond to the designed indicator, then this item will be evaluated. The experts found that the test items corresponded to the domain sample behavior of visual-spatial ability. However, some suggestions were given as to how items should be represented, such as clarifying rotation figures to reduce errors. On the basis of their suggestions, the rotation figures in the block rotation dimension were evaluated. The assessment also involves public opinion through legibility test, which is performed to determine whether target users can work on and solve the test items and understand the test instructions.

The next process is test trial. The data from trials will serve as the basis for measuring the psychometric aspects of tests, such as item analysis validity, reliability, and norm creation. Trials were conducted among groups of 10th grade high school students in Jakarta. The trials were conducted in the classroom and lasted approximately 35 minutes.

The next procedure is quantitative and qualitative analysis of the psychometric aspects of the test to determine the best items obtained by item pooling. A total of 20 items were selected as the final items of the visual-spatial test, which will be tested for reliability and validity to complement the psychometric aspects. Norms were also established to give meaning to the test scores.

Statistical Analysis

The constructed visual-spatial ability test based on CHC theory in this study uses several criteria: the test must contain items that can discriminate individuals on their visual-spatial ability, has distractors that perform well and are spread evenly among false options, is reliable, is valid, and has test scores that can be meaningfully interpreted.

Statistical procedures are performed to analyze the test results. The results are used to determine whether or not this research achieved its goals. The following criteria for each statistical procedure need to be met.

- Item analysis

Item analysis is performed on quantitative and qualitative aspects. The quantitative aspect is determined based on three indices: discrimination index, difficulty index, and distractor power

index. The qualitative aspect is considered based on form, content, and correspondence toward item design.

- **Item difficulty analysis**

Item difficulty index is determined based on the proportion of respondents who answered each item correctly (Anastasi & Urbina, 1997). Item difficulty index ranges from 0 to 1, where a great index means that the item is easy. Item difficulty index is divided into five categories that are used to analyze item difficulty by measuring the proportion of respondents who correctly answered an item compared with the total respondents on every test item (Anastasi & Urbina, 1997). Item difficulty is used to determine whether an item has an appropriate difficulty level to determine its complexity. Item difficulty index is also used to determine item order placement toward its difficulty.

- **Item discrimination analysis**

Item discrimination analysis measures item performance on differentiating respondent's ability toward the objective construct. Corrected item-total correlation (cr_{it}) is used to perform item discrimination analysis (Ebel & Frisbie, 1991). Corrected item-total correlation is performed by correlating each item score against the total score excluding targeted item in the total score. This result is compared with the criteria (0.2) given by Nunnally and Bernstein (1994). An item with index discrimination below that value will be revised or eliminated.

- **Distractor power analysis**

Distractor power analysis of responses is an advanced process in the construction of the visual-spatial test. This analysis aims to determine the functionality of the distractor's answer option in outwitting individuals who do not have good visual-spatial skills. An individual with good visual-spatial skills will be able to answer correctly, whereas an individual without good visual-spatial skills will be deceived into selecting the distractor's answer option (Friedenberg, 1995). Distractor power analysis was performed by comparing the ratio of expected distractor power and actual distractor power.

- **Reliability**

This research used Cronbach's alpha analysis for reliability analysis. Cronbach's alpha coefficient is a reliability testing method based on the consistency of responses with each test item and involves comparing the score of items with the total score of the rest of the test items (Anastasi & Urbina, 1997). This test is designed to measure one construct. Therefore, every item on this test should contain only that one construct. If an item does not contain the construct being measured, then determining the measurement result will be difficult, and the research goals will not be attained. According to Cohen and Swerdlik (2009), criteria that indicate the reliability of a test are categorized into three categories; Category A for reliability index ≥ 0.9 , Category B for reliability index ≥ 0.8 , and Category C for reliability index $0.65 \leq \alpha \leq 0.70$. Category C is the lowest criteria for a test to be categorized as a reliable test.

- **Validity**

The validity test of the visual-spatial test is conducted to check the validity of the constructs. The validity is checked by verifying the correlations with other tests. The validation tool used in this study is Tes Inteligensi Kolektip Indonesia-Menengah (TIKI-M) subtest 7 (Forming Shapes) and subtest 2 (Combined Parts). TIKI-M subtests 2 and 7 were selected based on several considerations, one of which is the validity of these subsets in measuring spatial ability constructs (Rachmawati & Andriani, 2014). In addition, TIKI-M tests are administered to 9th grade junior high school students up to 12th grade senior high school students, thereby indicating that the limit of measurement test ability is adjusted to junior and senior high school students

(Rachmawati & Andriani, 2014). This limit matches the constructed test, which also targets junior and senior high school students. The characteristics of these tests are equivalent to those of the constructed test. Hence, the validation process will use TIKI-M subtests 2 and 7.

Results

Item Analysis

The item analysis process involves item difficulty, item discrimination, and distractor power analyses. Both quantitative and qualitative item analyses were performed to examine the content, design, and item representation with the theoretical framework.

Item Difficulty Analysis

The table shows that item difficulty index is divided into four categories: the “very hard” category, which contains no test items; the “very easy” category, which contains 9 items; the “easy” category, which contains 9 items; the “medium” category, which contains 8 items; and the “hard” category, which contains 4 items. Some items have an uncategorized difficulty index. For example, 15 items are designed to have high difficulty, but results showed that the degree of difficulty of these items fall under the “very easy” category. Overall, these test items have a high item difficulty index, which means that the test items are relatively easy for participants to answer.

The result of item difficulty analysis is shown in Table 1.

Table 1
Item difficulty index

Difficulty Index Range	Category	Item Number	Total Amount
1.0–0.8	Very Easy	1,2,3,4,7,9,15,16,19	9
0.8–0.6	Easy	23,13,10,11,17,5,8,18,29	9
0.6–0.4	Medium	20,30,25,6,22,28,12,21	8
0.4–0.2	Hard	26,14,27,24	4
0.2–0.0	Very Hard	-	-
	TOTAL		30

Item Discrimination Analysis

According to statistical analysis, item discrimination index was performed by measuring the item score against the total score on the rest of the test items. The correlation coefficient of each item was then compared according to the criteria of Nunally and Bernstein (1994). On the basis of that criteria, two categories were created for the item discrimination index: good discrimination ability, which refers to items with a discrimination index above 0.2, and poor discrimination ability, which refers to items with a discrimination index below criteria 0.2.

Table 2 shows 21 test items that are categorized as “good discrimination ability,” which means these items can differentiate people who have or do not have the ability to answer correctly. Nine items are categorized as “poor discrimination ability,” which means they have insufficient capability to differentiate people’s ability and are thus evaluated or deleted.

The result of corrected item-total correlation measure is shown in the table below.

Table 2
Discrimination item index

Indicator cr_{it}	Description	Item Number	Amount
>0.2	Good discrimination ability	2,3,4,5,7,8,10,13,17,18,19, 20,21,22,23,25,26,27,28,29,30	21 items
<0.2	Poor discrimination ability	1,6,9,11,12,14,15,16,24	9 items
TOTAL			30 items

Distractor Power Analysis

Distractor power analysis result is divided into two categories: functioned and not functioned. An item is categorized as “functioned” if actual distractor power on every distractor and expected distractor power is equivalent. Table 3 shows that 21 items belong to the “functioned” category, which means each distractor on those items can optimally outwit the respondents without good visual–spatial skills. The items in the “not functioned” category will be evaluated or deleted from this test.

The results of distractor power analysis and its functionality are shown below.

Table 3
Distractor power analysis

Distractor Power	Item Number	Amount
Functioned	1,2,3,5,7,8,9,10,11,15,16, 17,18,21,22,23,26,27,28,29,30	21 items
Not Functioned	4,6,12,13,14,19,20,24,25	9 items
TOTAL		30 items

Integrative Item Analysis

Various item analysis results (quantitative and qualitative) served as the basis for integrative analysis. Integrative analysis was performed to choose the 20 best items from the available item pooling. This analysis also considered the proportionate representation of items in each dimension. Through this analysis, 15 items were eliminated because they did not meet the criteria for good items (quantitative and qualitative). The selected 20 items were tested for reliability and validity to measure the psychometric aspects of this test.

Reliability Test Results

Reliability test results of the Cronbach's alpha test showed that the visual–spatial test has a reliability coefficient of 0.714 ($M = 12.94$, $SD = 3.51$). This result means that from the 100% obtained observed score, 71% true variance and 29% error variance come from content sampling error and content heterogeneity error. The limit that Cohen and Swerdlik (2009) set stated that the reliability test limit is $0.65 \leq \alpha \leq 0.70$. Thus, this visual–spatial test has internal consistency in measuring one particular construct.

Validity Test Results

Validity test results showed a significant correlation between the visual–spatial test with TIKI-M subtest 7 score ($r = 0.337$, $p <.01$) and subtest 2 score ($r = 0.227$, $p <.05$.). Thus, the visual–spatial test is valid for measuring the spatial ability constructs because it has a significantly positive correlation with the TIKI-M subtest Combined Parts and Forming Shapes.

Norms

Norms creation begins with a normality data test using the Kolmogorov–Smirnov test. Normality test results showed that the data are not normally distributed ($p < 0.05$). Therefore, the results must be considered in norms creation. Norms creation does not use normalized standard score method and transforms the data to be normally distributed given the inadequate number of respondents. In addition, results show that the data are not normally distributed. Rather, the data skewed negatively, thereby indicating the homogeneity of the sample in terms of measurable ability, where the study sample has an average ability above the normal data rate. Norms creation with a normalized standard score is performed only with a large number of representative samples and if the deviation from normality comes from the test and not because of the characteristics of the sample or other factors in measurable behavior (Anastasi & Urbina, 1997). Therefore, the method used in norms creation is scaled score linear transformation.

The process of norms creation begins with the conversion of raw score to z-score. Z-score is a standard score test. The visual–spatial test has a mean value of 12.93 and a standard deviation of 3.51. To obtain the scaled score, the z-score is changed by multiplying the standard deviation and adding the new mean. The new deviation standard used is 3, and the new mean used is 10. The result of the change is shown in the table below.

Table 4
Visual–spatial test norms

Scaled Score (M = 10, SD = 3)	Raw Score	Scaled Score (M = 10, SD = 3)	Raw Score
16	20	7	9–10
15	19	6	8
14	18	5	7
13	16–17	4	6
12	15	3	5
11	14	2	3–4
10	13	1	2
9	12	0	0–1
8	11		

Discussion

This research aims to construct a visual–spatial ability intelligence test for 10th grade senior high school students. Reliability tests, validity, item analysis, and norms creation are part of the process of constructing new measuring instruments. The test results show that the visual–spatial test has good internal consistency and that its items are homogeneous in measuring one particular construct. The results of the validity test indicate that the visual–spatial test can measure the visual–spatial ability construct, as proven by a significantly positive correlation with valid validation tools that measure the visual–spatial ability construct. Item analysis proved that the level of difficulty of the measurement items of the visual–spatial test is in accordance with the optimum performance measurements of individuals and that the test has the discrimination

power to distinguish individuals with high visual-spatial ability and individuals with low visual-spatial ability. The item analysis results also indicate that the distractor answer option on each item of this test can function optimally in deception. On the basis of integrative item analysis, 20 items of the visual-spatial test were selected. Next, a group of norms is set using standard score with 10 for the mean and 3 for the standard deviation.

The test results on the psychometric aspects of the visual-spatial test indicate that the goal of the test was achieved. Psychometric test results show that the visual-spatial test can measure visual-spatial ability consistently where the items work properly. Significantly positive correlations with TIKI-M subtests 2 and 7 prove that the visual-spatial test measures the basic ideas of behavior similar to TIKI-M subtests 2 and 7. This valid result indicates that the test items can provoke a mental simulation response in the participants. Participants are required to perform mental simulation to be able to answer the test items correctly. Mental simulation is an important aspect that is emphasized in this test (Schrank et al., 2014). Schrank et al. (2014) explained that the implications in a test are derived through the process of detecting stimulus objects, performing mental simulations of stimulus objects, and matching with answer choices. The item shapes and the task or difficulty of the visual-spatial test are adjusted to suit senior high school students, consistent with Ramful et al. (2016). Ramful et al. (2016) stated that items can be made relevant to the respondents if they conform to the school curriculum. In the visual-spatial test, the relevance form is a special rotation angle that is familiar to senior high school students.

The integrative analysis results of item test and final item selection show that some items do not have good discrimination index but are still included in the final test item. This choice was made in reference to Ebel and Frisbie (1991), who said that the selection of available items must be prioritized by referring to the correspondence of the behavioral content. Therefore, the current study represents items according to the measured behavioral content.

The results of the visual-spatial test are not perfect. The level of reliability obtained by the visual-spatial test shows that the test is consistent. However, on the basis of Cohen and Swerdlik (2009), the value of the test coefficient is included in Category C or the lowest limit of test reliability. According to Anastasi and Urbina (1997), some aspects affect the value of reliability, one of which is variability. The reliability coefficient is highly dependent on the variability of the test sample. Similar to the correlation coefficient in general, variations in individual differences greatly affect the value of the coefficient of reliability. Test groups that are too homogeneous do not have significant differences in measurable ability. The effect is the occurrence of restricted range conditions, which result in low correlation. The participants in this research were obtained from three school with similar admission passing grade (Dinas Pendidikan Pemerintah Provinsi DKI Jakarta, 2015). These three schools have an average passing grade of 34. This finding is reinforced by research findings showing that the distributed data are negatively skewed, which means that participants gather on one side of the high score distribution. This characteristic indicates that participants are sampled homogeneously where the average participant has a high visual-spatial ability. Given the homogeneity of the sample, the range is restricted because the study sample has no variation that covers all possible levels of visual-spatial ability.

School selection is performed without considering the level of school quality. The process of selecting school samples does not use quality strata such as the passing grade of new students.

This limitation might impact the narrow reliability coefficient showed in the analysis.. Different results may be obtained if the test sample is more varied, thereby ensuring that the measuring instrument is applied to a sample that represents the population as a whole. High sample variations are expected to provide accurate test reliability values.

In addition to the problem of variability, the coefficient of reliability can also be affected by the error variance that occurs. Reliability testing using Cronbach's alpha results in variance errors, such as content sampling and heterogeneity errors. These errors occur because some items in the test do not collect a representative sample of visual-spatial ability skills. According to Ramful et al. (2016), this condition occurs because the items that are theoretically expected to measure the mental simulation process do not always correspond to what the participants do. The items are created using another simulation process depending on the participants viewing the presented stimulus. On this basis, Ramful et al. (2006) suggested clarifying the characteristics of visual-spatial abilities because this construct specification is still lacking in scientific research.

Homogeneous sample conditions also affect the norm creation process. The sample data show that the average sample score is fairly high and collected on one continuum. As a result, the sample does not meet the normal distribution. By contrast, in creating the norm, the comparison group meets the criteria of the standard group. In creating the group of norms, the sample of groups used for comparisons must be large enough and ensure population representation (Anastasi & Urbina, 1997). The population representation's characteristic is the variation ability of the represented Indonesian students. Therefore, improvements need to be made, especially in terms of the representation of sample measurement test samples to obtain accurate results.

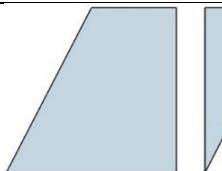
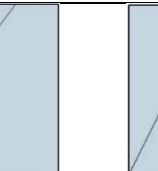
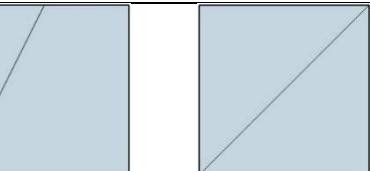
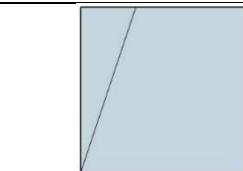
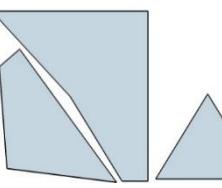
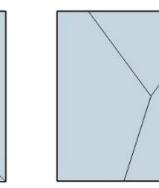
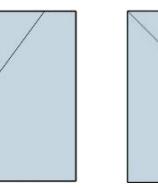
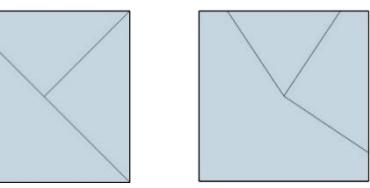
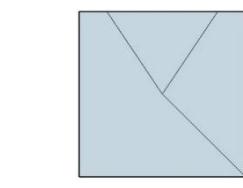
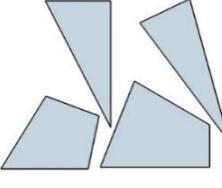
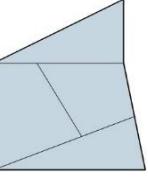
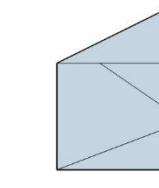
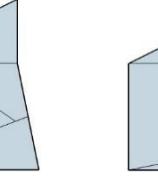
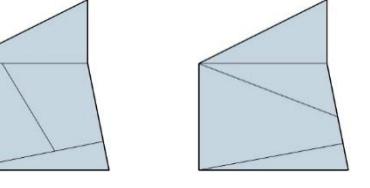
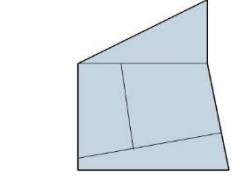
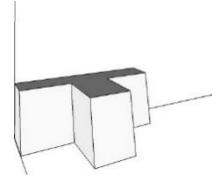
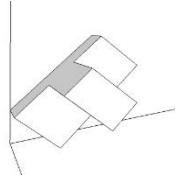
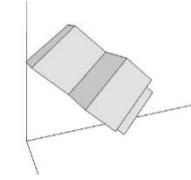
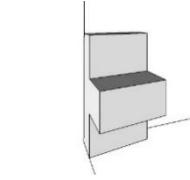
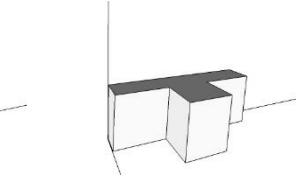
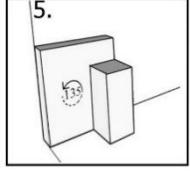
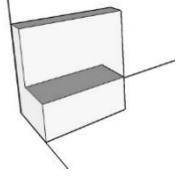
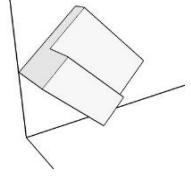
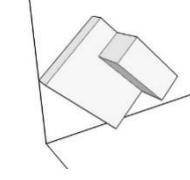
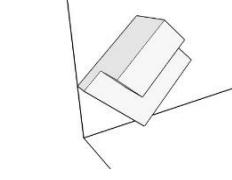
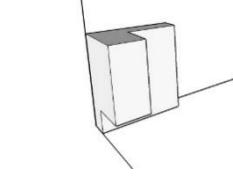
The limitation of this research is the lack of proof that the test is able to predict the success of individual learning in school. Roth et al. (2015) explained that intelligence can be a predictor of individual success or learning performance. On this basis, intelligence measurement should be able to predict individual success or learning performance. However, in the construction of this test, the relevant issue cannot be confirmed because the validity criteria test was not conducted using the criteria of student learning success. This requirement can be met by performing the validity criteria test using the criteria of student achievement.

Appendix

Table 1.
Construction structure of visual-spatial test items

Construct	Definition	Operational Definition	Dimension	*Characteristic
Visual-Spatial Ability	The ability to perceive a complex pattern and perform mental simulations of how these patterns are transformed (e.g., rotated, resized, and partially displayed) (Schneider & McGrew, 2012).	Able to perceive stimulus object and perform mental simulation as per item in each item against stimulus object*; match with answer choices (Schrank, McGrew, & Mather, 2014)	Space Relation: The ability to identify two-dimensional fractions into a form (Schrank, McGrew, & Mather, 2014) Block Rotation: The ability to perform mental rotate simulation on a three-dimensional shape (Schrank, McGrew, & Mather, 2014)	Space Relation: The simulation of the unification fractional form of the given question into one part remaining intact mentally. Block Rotation: The simulation of the space rotation given based on the given instruction and the result of the mental rotation.

Table 2.
Examples of visual-spatial tool item

Model	Variation	Question	A	B	C	D	E		
Space	Two pieces of shape fractions								
Relation	Three pieces of shape fractions								
	Four pieces of shape fractions								
Block	Rotation Angle $\leq 90^\circ$								
Rotation	Rotation Angle $> 90^\circ$	5.							

References

- Alnastasi, A., & Urbina, S. (1997). *Psychological Testing, (7th Ed)*. Indiana: Prentice-Hall.
- Benson, E. (2003). Intelligent intelligence testing: Psychologist are broadening the concept of intelligence and how to test it. *Monitor on Psychology*, 34, 2, 48.
- Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytic studies. Cambridge University Press.
- Cohen, R. J., & Swerdlik, M. E. (2009). *Psychological testing and assessment: An introduction to tests and measurements (7th ed.)*. New York, NY: McGraw-Hill.
- Dewi, F., & Nuryono, W. (2014). Survey tentang hambatan-hambatan selama proses peminatan (dalam konteks BK) berdasarkan kurikulum 2013 bagi siswa di SMA Negeri Se-kota Surabaya. *Jurnal BK UNESA*, 1-10.
- Dinas Pendidikan Pemerintah Provinsi DKI Jakarta. (2015, Juli 1). Penerimaan Peserta Didik Baru: Passing Grade. Retrieved from Penerimaan Peserta Didik Baru: <http://sma.ppdbsdki.org/passinggrade/index.aspx>
- DiPerna, J. C. (2006). Academic enablers and student achievement: Implications for assessment and intervention services in the schools. *Psychology in the Schools*, 43(1), 7-17.
- DiPerna, J. C. (2008). Academic Enablers: Assessment and Intervention Considerations.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement (5th ed)*. New Delhi: Prentice Hall
- Fan, X., & Chen, M. (2001). Parental involvement and student's academic achievement: A meta-analysis. *Educational psychology review*, 13(1), 1-22.
- Friedenberg, L. (1995). *Psychological testing. Design, analysis, and use*. Boston: Allyn and Bacon
- Johnson, J. F., Barron, L. G., Rose, M. R., & Carretta, T. R. (2017). Validity of Spatial Ability Tests for Selection into STEM (Science, Technology, Engineering, and Math) Career Fields: The Example of Military Aviation. In Visual-spatial Ability in STEM Education (pp. 11-34). Springer International Publishing.
- Kaufman, A. S., & Lichtenberger, E. O. (2006). Assessing adolescent and adult intelligence. John Wiley & Sons.
- Kementerian Pendidikan dan Kebudayaan. (2013). Pedoman Peminatan Peserta Didik. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- Kementerian Pendidikan dan Kebudayaan. (2014, Januari 14). kemdikbud.go.id. Retrieved from Kementerian Pendidikan dan Kebudayaan: <https://kemdikbud.go.id/kemdikbud/dokumen/Paparan%20Wamendik.pdf>
- Kementerian Pendidikan dan Kebudayaan. (2014). Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 64 Tahun 2014 tentang Peminatan pada Pendidikan Menengah. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- McGrew, K. S., & Wendling, B. J. (2010). Cattell–Horn–Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools*, 47(7), 651-675.
- Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.)*. New York: McGraw-Hill.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin*, 135(2), 322.
- Rachmawati, F. A., & Andriani, F. (2014). Confirmatory Factor Analysis Tes Inteligensi Kolektif Indonesia Tingkat Menengah (TIKI-M). *Jurnal Psikologi Pendidikan dan Perkembangan*, 3, 69-76.
- Ramful, A., Lowrie, T., & Logan, T. (2016). Measurement of spatial ability: Construction and validation of the spatial reasoning instrument for middle school students. *Journal of Psychoeducational Assessment*, DOI: 10.1177/0734282916659207
- Rozi, M., & Irawan, M. (2015). Sistem pendukung keputusan dalam memilih jurusan SMA menggunakan model Yager. *Jurnal Sains dan Seni ITS*, 7-12.
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118-137.
- Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of personnel evaluation in education*, 11(1), 57-67.
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. Dalam D. P. Flanagan, & P. L. Harrison, *Contemporary intellectual assesment: Theories, tests, and issues 3rd Edition* (hal. 99-144). New York: Guilford Press.
- Schrank, F. A., McGrew, K. S., & Mather, N. (2014). Woodcock-Johnson IV. Rolling Meadows, IL: Riverside
- Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research*, 4(1), 27-41.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of educational research*, 75(3), 417-453.
- Uttal, D. H., & Cohen, C. A. (2012). 4 Spatial Thinking and STEM Education: When, Why, and How?. *Psychology of Learning and Motivation-Advances in Research and Theory*, 57, 147.
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817.
- Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, 99(2), 397.
- Yayasan Dharma Bermakna. (2015). *Programs and Activities: Research*. Retrieved from Yayasan Dharma Bermakna: http://ydb.or.id/home/index.php?option=com_content&view=article&id=120&Itemid=1670