Psychometric Properties of Processing Speed Ability Test: a Pilot Project

Fitri Andriani
Faculty of Psychology Universitas Airlangga
Surabaya-Indonesia
fitri.andriani@psikologi.unair.ac.id

Cholichul Hadi
Faculty of Psychology Universitas Airlangga
Surabaya-Indonesia
cholichul.hadi@psikologi.unair.ac.id

Urip Purwono
Faculty of Psychology Universitas Padjajaran
Bandung-Indonesia
urip.purwono@gmail.com

Siti Sulasmi
Faculty of Business Economics Universitas
Airlangga Surabaya-Indonesia
Sulasmi_m@feb.unair.ac.id

Abstract

This study investigates the psychometric properties of a Cattel-Horn-Carroll theory-based processing speed ability test. According to the theory, processing speed ability, which has three narrow abilities (i.e., perceptual speed, number facility, and rate of test taking) supports and has a significant impact on general intelligence (about 0.7). A trial was conducted involving 137 people to test the quality of 299 composed items. Item selection and test reliability estimation were based on data analysis using ITEMAN. The result shows that the tests under investigation have sufficient psychometric properties and adequate reliability.

Keywords: *Processing speed ability, Intelligence, CHC Theory*

Received 14 March 2019/Accepted 25 May 2019 ©JEHCP All rights reserved

Introduction

Individual differences are an important subject matter in psychology. Understanding individual differences would provide information about the uniqueness of an individual as compared to others (Marnat, 2003). This is supported by the statement that to categorize the learning process also needs to consider individual differences, both qualitatively and quantitatively (Hickendor, Edelsbrunner, Mcmullen, Schneider, & Trezise, 2018). Individuals might differ in term of physical or psychology attributes. Physical differences include body height, complexion, blood type, and other physical qualities which characterize an individual. Psychological differences are generally distinguishable into two kinds of attributes, namely cognitive attributes and non-cognitive attributes. Cognitive differences may include intelligence, numeracy, memory, problem-solving skill, etc. Meanwhile, non-cognitive differences can comprise personality maturity, self-confidence, cooperation skill, adaptability, and so on. It is important to identify individual differences as they indicate the distinguishing qualities of an individual.

This research pertains to one particular psychological difference, which is intelligence. To this day, intelligence is still of great interest for researchers to study (Gottfredson & Saklofske, 2009). Researches on intelligence have had an extensive impact on the development of the community. Scientists believe that intelligence holds a vital role in human behavior. It is assumed to be able to predict individuals' academic achievement (Caemmerer, Maddocks, Keith, & Reynolds, 2018; Lohman & Gambrell, 2012; Lohman, Korb, Lakin, Korb, & Lakin, 2008; Naglieri & Ford, 2003; Sukadji, 1998; Tarbetsky, Collie, & Martin, 2016) and success in career (Flanagan & Mcgrew, 1998; Gottfredson & Saklofske, 2009; Sukadji, 1998), as well as its critical role for human survival (Anastasi & Urbina, 1997; Sukadji, 1998). Owing to its great significance, it is important to examine and measure the intelligence capacity of an individual. Intelligence in this research specifically pertains to cognitive speed processing or also known cognitive speediness. According to Carrol, cognitive speed processing is on the Stratum II in the Three-Stratum Theory of Cognitive Abilities (assuming there are three stratums in human cognitive structure, namely Stratum I, Stratum II, and Stratum III). There is still a limited number of tests to measure speed processing ability. The existing tests also focus more on measuring precision. Therefore, it is necessary to develop a new test specifically tailored to measure speed processing ability.

Accessing one's intelligence or other psychological attributes requires assessments which might involve a test (AERA, APA, & NCME, 2014). The test is a tool to collect data in assessment. Data collection can affect decision making. An inappropriate decision might be due to inaccurate data (Sukadji, 1998). A test is an objective and standardized instrument to measure a sample of behavior (AERA et al., 2014; Anastasi & Urbina, 1997). It has a significant role in determining a psychological decision in which the life of an individual, a group, or a community might be at stake. Hence, quality evaluation of a test should be carried out prior to its usage. Evaluation can be conducted during the initial development of the test, through item analysis, and through the estimation of its reliability (AERA et al., 2014; Marnat, 2003). This research evaluated the speed processing ability test in those said areas.

This study will discuss the psychometric properties of a test designed to measure speed processing ability, which was constructed based on the Three-Stratum Theory of Cognitive Abilities, or also known as the Cattel-Horn-Carrol (CHC) Theory. The psychometric properties under investigation include the item and overall test qualities.

Processing Speed Ability in the CHC Theory

This research focuses on constructing a test to measure speed processing ability (some might refer to it as cognitive speed processing or cognitive speediness). The definition of speed processing ability used here refers to the CHC theory. The CHC Theory integrates the Cattel and Horn's Gf-Gc Theory and Carrol's Three-Stratum Theory to explain general intelligence more comprehensively (Caemmerer et al., 2018). This integrative model arguably provides the best description of the structure of human mental abilities to its strong empirical foundation. Acknowledging the opinion of Carrol, McGrew and Flanagan recommend the integration of Cattell-Horn's Theory and that of Carroll (which was later known as the CHC Theory). It is a taxonomy combining findings of many factor analysis studies on intelligence. The CHC theory has an extensive implication on the measurement of intelligence (Flanagan & Harrison, 2005; Gregory, 2011; Mcgrew, 2009; Mcgrew & Flanagan, 1998; Newton & McGrew, 2010). According to this theory, intelligence comprises three hierarchical-structured abilities, namely pervasive, broad, and narrow ability (Caemmerer et al., 2018; Newton & McGrew, 2010). Carroll proposed no less than 69 narrow abilities, as shown in Figure 1. Meanwhile, McGrew revisited this in a study and found at least 59 narrow abilities.

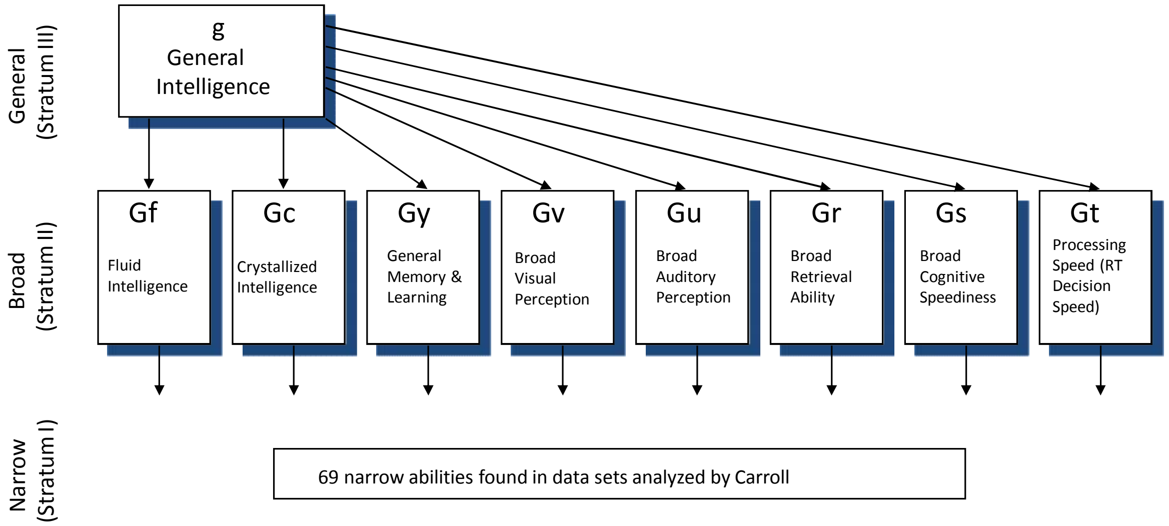
General (Stratum III)
g
General Intelligence
Broad (Stratum II)
Gf
Fluid Intelligence
Gc
Crystallized Intelligence
Gy
General Memory & Learning
Gv
Broad Visual Perception
Gu
Broad Auditory Perception
Gr
Broad Retrieval Ability
Gs
Broad Cognitive Speediness
Gt
Processing Speed (RT Decision Speed)
Narrow (Stratum I)
69 narrow abilities found in data sets analyzed by Carroll

Figure 1. Carroll's Three-Stratum Theory of Cognitive Abilities (1993)
Source: McGrew & Flanagan (1998).

Cognitive processing speed (Gs) refers to the speed in doing a sustainable learning or automatic cognitive process, especially when a high level of attention and concentration are required. Speed is thought to reflect the overall efficiency of the brain to register and process information (Tourva, Spanoudis, & Demetriou, 2016). For instance, the ability to complete simple mathematical operation quickly indicates a high level of speed processing ability. Ability to distinguish two words is also indicative of it. Speed processing ability has three narrow abilities, including perceptual speed, number facility, and rate of test taking. A detailed explanation of each narrow ability is shown in Table 1.

Table 1

Narrow Abilities of Cognitive Processing Speed

Ability	Description
Cognitive Processing Speed (Gs)	Ability to do fairly easy and familiar cognitive tasks, particularly those requiring a high degree of mental efficiency (e.g., concentration and attention), in an automatic and fluent fashion. It also refers to the speed in executing learned, basic cognitive processes automatically.
Perceptual speed (P)	Ability to quickly and accurately find, compare (spotting visual similarities and differences), and identify visual elements that are presented separately and side by side. Recent studies demonstrate that P is defined by four facets: (1) Pattern recognition (Ppr), is the ability to recognize simple visual patterns swiftly; (2) Scanning (Ps), is the ability to scan, differentiate, and seek visual stimulation; (3) Memory (Pm), is the ability to execute visual-perceptual speed tasks which directly requiresignificant short-term memory capacity; and (4) Complex (Pc), the ability to complete visual pattern recognition tasks which enforce additional cognitive demand, such as spatial visualization, estimation and interpolation, and increasing the load of memory span.
Number facility (N)	Ability to quickly execute basic arithmetical operation (such as addition, subtraction, multiplication, division), and to accurately and quickly manipulate numbers. N does not include comprehension or organization of mathematical problems and is not the main component of quantitative reason or higher mathematical ability.
Rate of test taking (R9)	Ability to quickly complete a relatively easy or well-learned (i.e., requiring a very simple decision) test. This ability is not content-related nor specific to any test stimuli.

Source: (Flanagan & Harrison, 2005; Flanagan & Mcgrew, 1998)

Cognitive processing speed has an important role in several aspects of intelligence. According to Nettelbec (1994), one commonly identified characteristic of intelligent behavior is mental speed (Mather & Wendling, 2005). Processing speed is the ability to quickly and smoothly complete simple tasks in a sustainable period. McGrew and Flanagan (1998) define processing speed as the ability to execute cognitive tasks automatically, particularly when under pressure to maintain focus and concentration, to swiftly process information with such limited resources and turn it into higher-order thinking. Perceptual speed is a narrow ability to process speed. Carroll describes it as an ability to find and compare visual symbols (Carroll, 1993). Highlighted by Mather, this ability is closely related to reading performance in elementary school, mathematical performance in elementary school and adulthood as well, and writing performance (Mather & Wendling, 2005). Thus, the ability to process symbols is associated with academic performance, particularly during elementary school. In a study investigating differences between good readers and those who could not read well, slower visual processing was found in students with a weak reading ability (Kruk & Willows, 2001).

Psychometric Properties of Cognitive Tests

Psychometric property is a psychometric aspect indicating the quality of an item or the overall scale. On item-level, according to Freeman, 1962 (Chadha, 2009), there are two important things to consider when determining the quality of an item, namely item discrimination, and item difficulty.

Item discrimination refers to the degree to which an item can accurately discriminate between test takers who have a higher level of the variable in question and those who are in the lower level (Domino & Domino, 2006). If an item has a good disclination power, more respondents with higher ability will be able to respond correctly to it, while less respondent with lower ability will be able to do so. It can be estimated through biserial correlation or point-biserial correlation. In tests with a dichotomous score, the use of point biserial correlation is more common (Urbina, 2004). The interpretation of point biserial correlation is similar to the correlation coefficient in general, in which coefficient ranges from -1.00 to +1.00, with a figure closer to 1 is better. Some experts suggest that a coefficient greater than .30 is deemed satisfactory.

An item difficulty pertains to the proportion of subjects that correctly respond to it. Bigger proportion implies that more people respond correctly, or the item was easy (Anastasi & Urbina, 1997; Hubley & Zumbo, 2013). Item difficulty is usually represented by the letter "p" for the proportion of correct. According to Dedrich, 1960, as quoted by Chada (Chadha, 2009), favorable index of item difficulty is between .35 to .85. Meanwhile, Shultz recommends difficulty index within the range of .25 to .80 (Shultz, Whitney, & Zickar, 2013).

In addition to psychometric properties of items, there are also properties of scale, namely validity, and reliability (Caviola, Primi, Chiesi, & Mammarella, 2017; Naglieri, 2013). Validity is the accuracy or precision of a scale to do its intended function. Validity is related to the objective of a measurement. There are at least five sources of validity evidence, including (1) test content, (2) response process, (3) internal structure, (4) relation to other variables, and (5) consequences of the test (AERA et al., 2014). These sources are complementary. The more evidence is obtained, the more confident the interpretation of a test score is.

Evidence-based on test content involves logical examination and evaluation against the test content to determine if it represents the variable which a test is intended to measure (AERA et al., 2014). Test content refers to the theme, diction, item format, task, or test questions, including administrative and scoring instructions. The procedures to obtain such evidence include logical analysis and expert review of test content's congruence with the defined construct, its relevance, importance, and item bias.

Evidence-based on response process pertains to what kind of responses it takes for test takers to complete the test (AERA et al., 2014). This type of evidence can be obtained by interviewing test takers to comprehend the reasons behind the respond to an item. Other feasible procedures are observation method and thinking aloud protocol.

Evidence-based on internal structure focuses on evaluating whether items correspond to the measured construct. It also involves confirmation of factors and exploration of test items to determine whether they are biased for different groups (AERA et al., 2014). Feasible procedures to obtain this kind of evidence include confirmatory factor analysis, cluster analysis, and inter-item correlational analysis. Confirmatory factor analysis is a factor analysis by confirming several empirical constructs assumed to be the factors of a latent construct. The objective of this analysis is to explain and describe by reducing the number of parameters. Reducing variables into higher level in confirmatory factor analysis is known as second-order factor analysis. In addition to reducing observed variables into some latent constructs, this kind of factor analysis also reduces the number of latent constructs into another latent construct. This method was carried out by Satre-Riba and friends when examining the psychometric properties of The Almost Perfect Scale-Revised (Sastre-riba, Pérez-albéniz, & Fonseca-pedrero,

2016) and Zhoc and Chen when they tested the validity of the Self Directed Learning Scale (Zhoc & Chen, 2016).

Evidence-based on relation to other variables can be obtained by correlating test score with another instrument which is expected to be correlated and with those which are expected not to be correlated (AERA et al., 2014). Some procedures to obtain such evidence are a correlation with external criteria, analysis of group differences, convergent and discriminant validity, multi-trait multi-method, and experimental or correlation studies regarding the construct in question. This method was carried out by Zhoc and Chen when testing the psychometric properties of the Self Directed Learning Scale (Zhoc & Chen, 2016).

Lastly, evidence-based on the consequences of a test is a new interpretation of validity, involving the evaluation of positive advantages and negative consequences of a test (AERA et al., 2014). One procedure to gather evidence from this source is to conduct a study about the realization of expected benefit and negative consequences of testing.

Meanwhile, reliability refers to the consistency of a measurement (Swerdlik & Cohen, 2005). The index of the ratio between variance of the true score and the total variance is called the reliability coefficient (Swerdlik & Cohen, 2005). There are several approaches to estimate reliability, including test-retest, parallel form, and internal consistency.

The test-retest approach is a method of reliability estimation by correlating two scores of the same sample from two separate administrations using the same test (AERA et al., 2014). It means that the test is delivered twice to a single group (Swerdlik & Cohen, 2005). This method requires only one form of a test, because the first (T1) and the second (T2) test are the same, making the measured construct exactly the same as well (Kline, 2005). It is more appropriate to estimate the reliability of a test which is intended to assess a relatively stable construct (AERA et al., 2014; Lee et al., 2017). When the construct being measured changes easily, subjects are likely to go through a learning process and acquire new ability during the interval between the first and the second testing. It will definitely affect the quality of the test (Swerdlik & Cohen, 2005). The estimation of test-retest reliability uses Pearson's formula of product moment correlation. Recent research, in 2017, this approach has been carried out in testing reliability against the scale of the School climate and student school identification measure (Lee et al., 2017).

The last approach is parallel form approach.This approach is carried out by giving two test kits that are parallel to the same group of subjects. This method is, however, difficult to employ because it requires two completely parallel forms of a test. Parallel tests should at least fulfill the following specification: the number of items must be the same, item format must be the same with regard to the involved trait content, item difficulty, and sampling adequacy; items should be evenly distributed, including the difficulty level; and both forms must be homogenous in term of the measured traits.

The internal consistency approach is conducted by administering one set of a test to a single group once (AERA et al., 2014). This method evaluates inter-item stability or consistency.The advantage of this method is that it allows the estimation of reliability using a single sample in only one test administration.Another advantage is that it can eliminate the effect of differences between the two measurements (AERA et al., 2014). Several methods to estimate reliability using this approach are available, including by splitting the test into two halves and then calculating the reliability estimate using a formula proposed by Cronbach known as the Cronbach's alpha formula (AERA et al., 2014; Zhoc & Chen, 2016), a formula by Kuder Richardson, or another one by Rulon and Flanagan. This approach has been carried out by Panggabean and Himawan in testing reliability against the scale of the Indonesian Teacher Competence Questionnaire (Panggabean & Himawan, 2016).

Steps of Test Development

Several general steps must be employed in developing a test so that the test result has an accountable procedural strength. The steps are detailed in the Standard for Educational and Psychological Testing (AERA, 1999; AERA et al., 2014), as follows: (1) Identification of the main usage of the obtained score (Standard 1.1). In this step, a test developer determines the usage of the scores obtained from the test under development. In this current research, scores will be used to diagnose an individual's processing speed. The second (2) step is to define the trait/domain to be measured (Standard 1.2 and 1.7).This step determines the definition of the domain in question, which can be done through several methods (e.g., literature review, content analysis, critical incidents, observation, expert judgments, and learning the instructional objectives).

The next step is (3) to develop a test specification (Standard 1.6, 1.18, and 3.3). Some things to consider in developing a specification is determining item proportion for each sub-domain and setting the item specification. The subsequent step is (4) item development and review (Standard 3.6). Item development is a step where items are pooled through item writing according to the predetermined specification. It is crucial to decide the most appropriate item format, to verify if the chosen format is suitable for the targeted test takers, select item writers (and also, if necessary, to train those writers), and eventually to construct items. This step is followed by a review (revisiting and rewriting the item pool). During this step, experts review the pooled items to evaluate whether they have met the test specification and are in the correct format. The experts' judgment should pertain to items' relevance to the measured domain, clarity, and simplicity.

The fifth (5) step is field testing (Standard 3.8). After items are revised as the result of a pilot testing, the next step is to run a field testing. The aim is to select items with good qualities based on several predetermined criteria (i.e., item disclination and difficulty). A field testing can also be preceded by a pilot study, in which a test is administered to a group of people to examine whether its items are understandable. Instead of item omission, items are revised in this step when necessary to ensure better understandability. Standard test completion time is also estimated during this step. Subsequently, the sixth (6) step is determining the scoring procedure (Standard 3.22 and 3.23). It is the step where an item scoring procedure should be clearly defined to increase scoring accuracy. This step is followed by (7) the construction of test administration and instruction (Standard 3.20 and 3.21). This includes designing administration procedure and test instruction so that testers can administer it according to the aim of the test developers, as written in the administration instruction.

The next step is (8) item analysis (Standard 3.9 and 3.10). This is the step before item selection. The result of this analysis determines whether an item will be included in the final form of the test. To do this, a set of criteria of what comprises good items is needed, such as adequate item discrimination, a particular degree of difficulty, and well-functioning distractors. Analysis can be conducted with the help of related software (e.g., ITEMAN). After this, the following steps are (9) reliability estimation (Standard 2.1 and 2.2) and (10) validation study (Standard 1.3), in which the test reliability is estimated, and studies are conducted to find evidence of the test validity from various sources as detailed in the STANDARD (AERA, 1999; AERA et al., 2014). Test validation is conducted concerning the underlying theory. The last step is (11) norm development (Standard 4.1 and 4.10), wherein norm and manual to administer the test are developed.

Method

This research employs a quantitative approach. The test development in this research followed the recommended procedure or steps in the STANDARD, as follows: identification the measurement objective, defining the trait/domain to be measure, development of test specification, item pooling and review, field testing, determining scoring procedure, item analysis, reliability estimation, and validation study (AERA, 1999; AERA et al., 2014). Participants of this research were university students who participated in a selection process of tutors for children at-risk of dropping out of school, held by the Department of Social Affairs of Surabaya City. This tutor recruitment involved students from several universities in Surabaya, with a total of 137 participants, comprising 105 females and 32 males. The instruments were used in this research, each representing narrow abilities of speed processing ability, namely perceptual speed, number facility, and rate of test taking. In total, there were 299 items in the pool, of which 100 items are in the perceptual speed subtest, another 100 in the number facility subtest, and the rest 99 items are in the rate of the test-taking subtest. Expert review before the field testing was conducted to obtain Content Validity Ratio (CVR) and Content Validity Index (CVI). After that, data was analyzed using ITEMAN software to calculate the statistics of each item and the overall scale. Item statistics included mean, standard deviation, item difficulty, and biserial correlation. Meanwhile, the scale statistics were mean of difficulties, mean of biserial correlations, reliability, and the standard error measurement (SEM).

Result

As the STANDARD suggests, items were reviewed following their construction. There were eleven experts asked to review the items in term of their relevance, clarity, and simplicity. The experts were asked to rate from 1 to 4, which were then used to calculate the CVR and CVI. Based on the calculation, the minimum obtained CVR was .83, and the minimum obtained CVI was .98. According to Lawshe (Lawshe, 1975), with a total of 11 experts, the minimum required CVR is .59. Thus, the CVR and CVI of the items in this research were adequate, meaning that the constructed items were relevant to the objective of measurement, clear and understandable for testees, and simple in expressing the measure attribute.

The subsequent step was to analyze items using ITEMAN software. Item analysis aimed to select items based on item statistics and to estimate reliability. As shown by the output on ITEMAN software, the reliability estimate was using Cronbach's Alpha formula. There were three separate narrow abilities, and each statistic was provided, comprising several items, several subjects, mean, standard deviation, variance, Alpha reliability, SEM, mean of item difficulties, mean of item-total correlations, and mean of biserial correlations. Table 2 shows the item statistics of each subtest. Meanwhile, Table 3, 4, and 5 provide the scale statistics of perceptual speed, number facility, rate of test-taking subtest, respectively.

Table 2

Item Statistics

Scale Statistics	Perceptual Speed	Number Facility	Rate of Test Taking
Proportion Correct	0.022-0.985	0.015-0.978	0.153-0.934
Biserial	-0.781-0.819	-0.425-0.758	0.169-0.906
Point Biserial	-0.241-0.521	-0.193-0.519	0.103-0.621

Table 2 describes the item statistics for perceptual speed, number facility, and rate of the test-taking subtest. Out of 100 items in the perceptual speed subtest, the proportion of correct ranged from .022 to .985, that of the number facility subtest ranged from -.425 to .758, while the proportion in the rate of test-taking subtest ranged from .169 to .906. Concerning the point biserial correlation, out of 99 items in the perceptual subtest, the coefficients varied between -.241 and .521, in the number facility subtest they varied between -.193 and .519, while the coefficients were between .103 and .621 in the rate of taking a subtest. Based on these results, 60 items were selected in each subtest.

Table 3

Scale Statistics of Perceptual Speed Subtest

Scale Statistics	Scale
Number of items	100
Number of examinees	137
Mean	30.657
Variance	36.853
Standard deviance	6.071
Skewness	-0.328
Kurtosis	1.623
Minimum	9.000
Maximum	47.000
Median	31.000
Cronbach's Alpha	0.796
Standard of error measurement	2.739
Mean of P	0.307
Mean of Item-Total Correlations	0.257
Mean of Biserial Correlations	0.389

Table 3 shows that perceptual speed subtest had Alpha reliability coefficient of .796. This figure is deemed adequate for a newly constructed test, implying that the test score is reliable. Mean of item difficulties was .307; demonstrating that some items had a fairly high level of difficulty. The mean of item-total correlation was .257, indicating a satisfactory correlation between each item and total score. It means that each item measured the same thing as the total score of the subtest. Also, the mean of biserial correlation was .389; meaning that items of the perceptual speed subtest had fairly good discriminative power.

Table 4

Scale Statistics of Number Facility Subtest

Scale Statistics	Scale
Number of items	100
Number of examinees	137
Mean	20.255
Variance	40.205

Scale Statistics	Scale
Standard deviance	6.341
Skewness	-0.855
Kurtosis	1.085
Minimum	0.000
Maximum	31.000
Median	21.000
Cronbach's Alpha	0.800
Standard of error measurement	2.839
Mean of P	0.203
Mean of Item-Total Correlations	0.262
Mean of Biserial Correlations	0.403

Table 4 illustrates that the number of facility subtest had Alpha reliability coefficient of .800. This figure is deemed satisfactory for a test, meaning that its score is reliable, especially when it is still in a pilot project.The mean of item difficulties (p) was .203; showing that the majority of items had a fairly high level of difficulty. Further, the mean of item-total correlation was .262, while the mean of biserial correlation was .403; which indicated a relatively strong association between each item and the sum score of all items, and that the items had fairly good discrimination.

Table 5

Scale Statistics of Rate of Test Taking Subtest

Scale Statistics	Scale
Number of items	99
Number of examinees	137
Mean	31.204
Variance	145.433
Standard deviance	12.060
Skewness	-0.417
Kurtosis	-0.176
Minimum	0.000
Maximum	54.000
Median	33.000
Cronbach's Alpha	0.923
Standard of error measurement	3.346
Mean of P	0.315
Mean of Item-Total Correlations	0.433
Mean of Biserial Correlations	0.574

As shown in Table 5, the rate of test-taking subtest had the Alpha reliability coefficient of .923. Reliability coefficient greater than .9 implies that the test score is reliable. The mean of item difficulties was .315, implying that some items had a fairly high level of difficulty. The mean of item-total correlation was .433, while that of biserial correlation was .574, which indicated that the items had good discrimination indices. The item discrimination index in this subtest was the highest, as compared to the number facility and perceptual sped subtest.

Discussion

The results showed that items had adequate psychometric properties as indicated by the proportion correct, biserial, and biserial points. Also, the three subtests compiled have quite good validity, which can be seen from the Content Validity Ratio and Content Validity Index. The reliability coefficient above 0.7 also proves the consistency of measurement.

Selection of "good" items, i.e., items with satisfactory psychometric properties, in the current research was based on their discrimination and difficulty. This is in accordance with the suggestion by Freeman (1962), as quoted by Chadha (Chadha, 2009). Table 2 provides the ranges of items statistics from the lowest to the highest value, which was used as the basis in selecting items. Table 3, 4, and 5 summarize the scale statistics of each narrow ability. The overall scale statistics should consider the Alpha coefficient, SEM, mean of item difficulties, mean of item-total correlation, and mean of biserial correlations.

In general, analysis of the 299 items tested on 137 participants yielded evidence that items had an adequate mean of discrimination power. This was indicated by the mean values of biserial correlations in every subtest, which were all greater than 0.3. Mean of item-total correlation and mean of the biserial coefficient are the correlation between item score and the total score. The high coefficient in these correlations indicates that items are analogous to what was measured by the total score. Shultz proposed value greater than 0.4 as an indication of good item disclination.

Meanwhile, the discrimination index between 0.3 and 0.39 is deemed satisfactory but needs further improvement. Items with discrimination power between 0.20 and 0.29 are considered unsatisfactory, and revision is required. Discrimination index smaller than 0.2 implies that an item should be omitted (Shultz et al., 2013). In this research, the biserial correlation coefficients in all three subtests were greater than 0.3, indicating that items could discriminate participants with high ability from those with lower ability.

Similarly, item difficulty indices in this research were considered acceptable, despite several difficult items. As mentioned before, item difficulty is represented by the symbol "p," abbreviation of "proportion of correct. According to Dedrich, 1960, as quoted by Chadha (Chadha, 2009), favorable item difficulty is between 0.35 and 0.85. In contrast, another expert recommended the value between 0.25 and 0.80 for item difficulty (Shultz et al., 2013). In more details, Shultz categorizes p greater than 0.8 as very easy item, p between 0.5 and 0.8 as easy item, p between 0.25 and 0.49 as difficult item, while p smaller than 0.25 is classified as the very difficult item (Shultz et al., 2013). In the current research, the mean of p statistics was ranging from .203 to .315, meaning the items were categorized as difficult (although, based on Table 2, some items were very easy).

Pertaining to distractors, items containing distractors (usually in a multiple-choice test) require distractor analysis to see whether the distractors are well-functioning. According to Urbina, the indication of an ideal distractor for multiple-choice items is when students with higher ability are not affected by it because they know the correct answer, while those with lower ability are affected because it seems correct for them (Urbina, 2004). In this research, the items did not contain any distractor, thus an analysis of distractor, despite being common to use as an indicator of a good item, was not conducted.

Further, the Alpha coefficients in all three subtests were found greater than 0.7, with one subtest (i.e., rate of test taking) had the coefficient of 0.9, providing evidence for them as reliable instruments. Referring to Aiken, those obtained coefficient are considered acceptable. Aiken suggests that reliability coefficient of 0.6 is adequate, although unsatisfactory (Aiken, 2003). In addition to reliability, SEM was also calculated. It indicates the magnitude of error in the measurement. SEM is obtained from a computation involving standard deviation and reliability. Lower SEM implies a more reliable test. According to Kelley (1927), SEM is useful to estimate discrepancy between individuals' true score and their observed score; between an observed score of one form of a test and observed score of its parallel form; as well as between true score and estimated true score (Crocker & Algina, 2008). This research found SEM value ranging from 2.739 to 3.346.

Using the result of item analysis as the basis, the next step was to select items which met the criteria. In the current pilot project, 60 items from each narrow ability were selected. Several total items were 180, with completion duration of four minutes for each subtest. In total, it takes fifteen minutes to complete this processing speed test.

To conclude, this newly developed test to measure processing speed ability, though the measurement of test taking, number facility, and perceptual speed, had satisfactory psychometric properties. Content validation conducted through expert judgment yielded a result that the constructed test was able to measure processing speed ability. Therefore, as a pilot project of a test development, which is important in addressing the dearth of test measuring fast thinking process, the current test is deemed fairly reliable for now. It is assumed to be able to predict students' success in recognizing symbols, reading, information processing, and more importantly, their intelligence. Research shows that intelligence is predicted by speed and accuracy (Borter, Troche, & Rammsayer, 2018).

The set of tests produced in this study are also a benefit of this study, especially the benefits for the field of psychological practice in providing alternative tests for measuring processing speed abilities. For the next step, further validation and standardization should be conducted through the collection of evidence from various sources as suggested in the STANDARD (AERA, 1999; AERA et al., 2014) including evidence based on test content, response processes, internal structure, internal structure, relation to other variables, and consequences of test. Validation and standardization can also be conducted through further studies on speed processing test.

Conclusion

This research concludes that the speed processing test had a fairly good quality. Validation based on test content through a calculation of CVR and CVI yielded an adequate result (i.e., CVR and CVI values were greater than the predetermined criterion). Additionally, item statistics indicated adequate qualities, as indicated by several criteria, including item discrimination and difficulty. Moreover, scale statistics also met the criteria for acceptable reliability, mean of item-total correlation, mean of biserial correlation, and mean of item difficulty. Based on its psychometric properties, this test could be used to measure speed processing. Further studies, however, are necessary to ensure its validity based on other sources.

A recommendation for test users is to be careful in employing this test because no norm is available yet to base interpretation of the score. Future researches should test the construct and criterion-related validity of the test to strengthen the evidence of its unified validity, as recommended by the STANDARD (AERA, 1999; AERA et al., 2014). Validation studies can be conducted through factor analysis, analysis of correlation with other tests, or by examining differences between two groups with evidently different level of processing speed ability. Another recommendation is to include more diverse participants in term of age in future field testing so that norm development can be feasible.

References

AERA, A. & N. (1999). *Standards: Educational And Psychological Testing*. Washington: American Educational Reseach Association.

AERA, APA, & NCME. (2014). *The Standards for Educational and Psychological Testing*.

Aiken, L. R. (2003). *Psychological: Testing And Assessment*. New York: Pearson Education Group, Inc.

Anastasi, A., & Urbina, S. (1997). *Psychological Testing* (7th Ed). New York: Prentice Hall.

Borter, N., Troche, S. J., & Rammsayer, T. H. (2018). Speed and Accuracy-Related Measures of An Intelligence Test are Differentially Predicted by The Speed and Accuracy Measures of A Cognitive Task. *Intelligence*, *71*(September), 1–7. https://doi.org/10.1016/j.intell.2018.09.001

Caemmerer, J. M., Maddocks, D. L. S., Keith, T. Z., & Reynolds, M. R. (2018). Effects of Cognitive Abilities on Child and Youth Academic Achievement : Evidence from the WISC-V and WIAT-III. *Intelligence*, *68*(February), 6–20. https://doi.org/10.1016/j.intell.2018.02.005

Carroll, J. B. (1993). *Human Cognitive Abilities A Survey of Factor-Analytic Studies*. New York: Cambridge.

Caviola, S., Primi, C., Chiesi, F., & Mammarella, I. C. (2017). Psychometric properties of the

Abbreviated Math Anxiety Scale (AMAS) in Italian primary school children. *Learning and Individual Differences*, *55*, 174–182. https://doi.org/10.1016/j.lindif.2017.03.006

Chadha, N. K. (2009). *Applied Psychometry*. New Delhi: Sage Publication. https://doi.org/10.4135/9788132108221

Crocker, L., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory*. *Cengage Learning*. Mason: Cengage Learning.

Domino, G., & Domino, M. L. (2006). *Psychological Testing: An Introduction* (2nd ed). Cambridge: Cambridge University Press. https://doi.org/10.4102/sajip.v35i1.807

Flanagan, D. P., & Harrison, P. L. (2005). *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (2nd Ed.). New York: The Guilford Press. https://doi.org/10.1080/02783190802201986

Flanagan, D. P., & Mcgrew, K. S. (1998). Interpreting Intelligence Tests from Contemporary Gf-Gc Theory : Joint Confirmatory Factor Analysis of the WJ-R and KAIT in a Non-White Sample, *36*(2), 151–182.

Gottfredson, L., & Saklofske, D. H. (2009). Intelligence : Foundations and Issues in Assessment. *Canadian Psychological Association*, *50*(3), 183–195. https://doi.org/10.1037/a0016641

Gregory, R. J. (2011). *Psychological Testing, History, Principles, and Applications*. Boston: Pearson Eucation, Inc.

Hickendor, M., Edelsbrunner, P. A., Mcmullen, J., Schneider, M., & Trezise, K. (2018). Informative Tools for Characterizing Individual Differences in Learning : Latent Class, Latent Profile, and Latent Transition Analysis. *Learning and Individual Differences*, *66*(October 2017), 4–15. https://doi.org/10.1016/j.lindif.2017.11.001

Hubley, A. M., & Zumbo, B. D. (2013). Psychometric Characteristics of Assessment Procedures: An Overview. In K. F. Geisinger (Ed.), *APA Handbook of Testing and Assessment in Psychology* (pp. 3–20). Washington: American Psychological Association.

Kruk, R. S., & Willows, D. M. (2001). Backward Pattern Masking of Familiar and Unfamiliar Materials in Disabled and Normal Readers. *Cognitive Neuropsychology*, *18*(1), 19–37.

Lawshe, C. (1975). A Quantitative Approach To Content Validity. *Personnel Psychology*, (1), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lee, E., Reynolds, K. J., Subasic, E., Bromhead, D., Lin, H., Marinov, V., & Smithson, M. (2017). Development of A Dual School Climate and School Identification Measure – Student (SCASIM-St). *Contemporary Educational Psychology*, *49*, 91–106. https://doi.org/10.1016/j.cedpsych.2017.01.003

Lohman, D. F., & Gambrell, J. L. (2012). Using Nonverbal Tests to Help Identify Academically Talented Children. *Journal of Psychoeducational Assessment*, *30*(1), 25–44. https://doi.org/10.1177/0734282911428194

Lohman, D. F., Korb, K. A., Lakin, J. M., Korb, K. A., & Lakin, J. M. (2008). Identifying Academically Gifted English- A Comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly*, *52*(4), 276–296. https://doi.org/10.1177/0016986208321808

Marnat, G. G. (2003). *Handbook Of Psychological Assessment* (4th Ed.). New Jersey: John Willey & Sons, Inc.

Mather, N., & Wendling, B. J. (2005). Linking Cognitive Assessment Results to Academic Interventions for Students with Learning Disabilities. In *Contemporary Intellectual Assesment:Theories* (pp. 269–294). New York: The Guilford Press.

Mcgrew, K. S. (2009). Editorial CHC Theory And The Human Cognitive Abilities Project : Standing On The Shoulders of The Giants of Psychometric Intelligence Research. *Intelligence*, *37*(1), 1–10. https://doi.org/10.1016/j.intell.2008.08.004

Mcgrew, K. S., & Flanagan, D. P. (1998). *The Intelligence Test Desk Reference (ITDR): Gf-Gc Cross-Battery Assessment*. Boston: Allyn & Bacon.

Naglieri, J. A. (2013). Psychological Assessment by School Psychologists: Opportunities and Challenges of a Changing Landscape. In K. F. Geisinger (Ed.), *APA Handbook of Testing and Assessment in Psychology* (pp. 3–20). Washington: American Psychological Association.

Naglieri, J. A., & Ford, D. Y. (2003). Addressing the Underrepresentation of Gifted Minority Children Using the Naglieri Nonverbal Ability Test (NNAT). *Gifted Child Quarterly*, *47*(2), 155–160. https://doi.org/10.1177/001698620304700206

Newton, J. H., & McGrew, K. (2010). Introduction to the special issue: Current Research in Cattel-Horn-Carrol Based Assessment. *Psychological in the School*, *47*(7).

Panggabean, M. S., & Himawan, K. K. (2016). The Development of Indonesian Teacher Competence Questionnaire. *Journal Of Educational, Health, and Community Psychology*, *5*(2), 1–15.

Sastre-riba, S., Pérez-albéniz, A., & Fonseca-pedrero, E. (2016). Assessing Perfectionism in Children and Adolescents : Psychometric Properties of The Almost Perfect Scale-Revised. *Learning and Individual Differences*, *49*, 386–392. https://doi.org/10.1016/j.lindif.2016.06.022

Shultz, K. S., Whitney, D. J., & Zickar, M. J. (2013). *Measurement Theory in Action* (2nd Ed). London: Routledge.

Sukadji, S. (1998). *Perkembangan Konsep, Teori, Dan Pengukuran Inteligensi*. Pidato pengukuhan Upacara Penerimaan Jabatan Guru Besar Ilmu Psikologi Fakultas Psikologi Universitas Indonesia.

Swerdlik, M. E., & Cohen, R. L. (2005). *Psychological Testing and Assessment: An Introduction to Test and Measurement* (6th ed.). New York: McGraw-Hill.

Tarbetsky, A. L., Collie, R. J., & Martin, A. J. (2016). The Role of Implicit Theories of Intelligence and Ability in Predicting Achievement for Indigenous (Aboriginal), Australian students.

Contemporary Educational Psychology, *47*, 61–71. https://doi.org/10.1016/j.cedpsych.2016.01.002

Tourva, A., Spanoudis, G., & Demetriou, A. (2016). Intelligence Cognitive Correlates of Developing Intelligence : The Contribution of Working Memory, Processing Speed, and Attention. *Intelligence*, *54*, 136–146. https://doi.org/10.1016/j.intell.2015.12.001

Urbina, S. (2004). *Essential of Psychological Testing*. (A. S. Kaufman & N. L. Kaufman, Eds.). New Jersey: John Willey & Sons, Inc.

Zhoc, K. C. H., & Chen, G. (2016). Reliability and Validity Evidence for the Self-Directed Learning Scale (SDLS). *Learning and Individual Differences*, *49*, 245–250. https://doi.org/10.1016/j.lindif.2016.06.013