Buletin Psikologi
Volume 30, Number 2, 2022: 348– 360
DOI: 10.22146/buletinpsikologi.74475

ISSN 0854 - 7106 (print)
ISSN 2528-5858 (Online)

Measuring Fluid Reasoning and Its Cultural Issues: A Review in the Indonesian Context

Hanif Akhtar*[1,2],
[1] *Doctoral School of Psychology, ELTE Eötvös Loránd University, Hungary.*
[2]*Faculty of Psychology Universitas Muhammadiyah Malang, Indonesia,*

Submitted 3 May 2022 **Accepted 19 September 2022** **Published 23 December 2022**

Abstract. Fluid reasoning (Gf) is a central component of human abilities and has been shown to be statistically indistinguishable from general cognitive ability. However, measuring fluid reasoning is not that easy since the operationalization of the construct itself is not conclusive. This situation is exacerbated if the measurements and comparisons are carried out in groups with different cultural and language abilities. The purpose of this paper is to provide a brief overview of the fluid reasoning construct as well as measurement practices that have been carried out in Indonesia. My review would mention the positioning of fluid reasoning in several intelligence theories, how researchers conceptualize fluid reasoning, how researchers measure fluid reasoning in a multicultural context and outline critical future directions for test development. In conclusion, I emphasize the importance of using a multi-processes figural fluid reasoning test as a proxy for a culture-fair intelligence test that can be used in cross-cultural assessments.

Keywords: fluid reasoning; individual differences; intelligence; multicultural

Introduction

Fluid reasoning or fluid intelligence (Gf) is one of the central constructs in all major theories of intelligence. Schneider and McGrew (2012) described fluid reasoning as the deliberate but flexible control of attention to solve novel problems that cannot be performed by relying exclusively on previously learned knowledge. It is linked to the problem-solving type of intelligence in a novel situation. This type of thinking includes identifying relations, concept formation, generating and testing hypotheses, drawing inferences, classification, problem-solving, generalization, understanding implications, extrapolating, and transforming information (McGrew, 2009; Newton & McGrew, 2010). Horn (1967) noted that it is formless and can "flow into" a large variety of cognitive activities. Thus, fluid reasoning is considered the cornerstone and plays a central role in human cognition.

Several researchers have suggested that fluid reasoning is an essential aspect of education in science and mathematics (Blankson & Blair, 2016; Green et al., 2017; Kanari & Millar, 2004; Schmidt, 2002), as well as reading and writing skills (Cormier et al., 2016; Floyd et al., 2008; Guerin et al., 2020; Motallebzadeh & Yazdi, 2016). Researchers and educators have emphasized how important it is to assess an individual's reasoning skills (Barkl et al., 2012) which could be considered an

*Address for correspondence: akhtar.hanif@ppk.elte.hu

excellent measure of *g* (Frick et al., 2010). Test of fluid reasoning is often used in high-stakes exams (e.g., personnel selection) since novel problem-solving is common in most activities. In an ideal way, measuring fluid reasoning should be grounded in psychological efficiency and is considered independent of education and acculturation (Horn & Cattell, 1966). However, in some cases, measurement is conducted in a multicultural setting with different language proficiency and educational opportunities.

Specifically for a multicultural country like Indonesia, measuring fluid reasoning is challenging. However, to my knowledge, no literature explicitly discusses the practice of fluid reasoning measurement, especially in Indonesia. Thus, this review attempted to bridge the methodological gap to generate new insight regarding the measure of fluid reasoning. This review examines the constructs of fluid reasoning and its measurement in a multicultural context. My review will mention several theories of intelligence, how researchers conceptualize fluid reasoning, and how researchers measure fluid reasoning. A proposal agenda for test development will be mentioned as well. I conclude by summarizing the important challenge of measuring fluid reasoning and providing a recommendation for research in the future.

Discussion

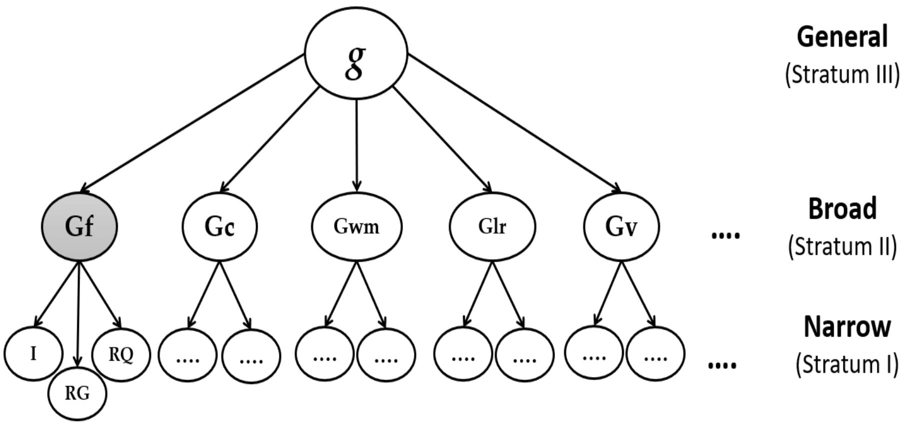
Fluid Reasoning in Intelligence Theory

Fluid reasoning was first proposed by Cattell (1963) along with the companion construct crystallized intelligence (Gc) to represent two distinct dimensions of Spearman (1904) General intelligence *(g)*. Fluid reasoning was described as the ability to solve novel problems using reasoning and was primarily biologically determined, while crystallized intelligence, in contrast, was a knowledge-based ability and was learned through education and experience (Cattell, 1983). This model was then expanded further by his student, Horn (1968), who found that there were more than two broad abilities. According to Cattell and Horn, intelligent behavior is best characterized by fluid reasoning.

In 1993, John Carroll published his book *Human cognitive abilities: A survey of factor-analytic studies*. Carroll (1993) summarized and reanalyzed factor-analytic studies of human cognitive abilities. The main strength of Carroll's meta-factor analysis was to provide an empirically based taxonomy of human cognitive ability in a single structured framework for the first time ever. The result of this work is an extensive theory of hierarchy called the "three-stratum model". The first stratum of his hierarchy is several narrow abilities. At the second stratum are eight abilities, and *g* is at the top of the structure as the third stratum. The Gf-Gc theory developed by Cattell and Horn (Horn & Noll, 1997) is closely linked to Carroll's model. The major differences between the three-stratum theory of Carroll and the Gf-Gc theory of Horn and Cattell are the lack of a general factor in the Cattell-Horn model since the nature of the *g* is determined by the composition of the test battery.

An integrated Cattell-Horn and Carroll model proposed by McGrew (1997) was the first attempt to create a single taxonomy (Figure 1). Both John Horn and John Carroll accepted the unified model proposed by McGrew and colleagues and thus became known as the Cattell-Horn-Carroll (CHC) theory. It is the most comprehensive and empirically supported psychometric theory of the structure of cognitive abilities to the present day (Flanagan & Dixon, 2014). Fluid reasoning forms a major part of the CHC structure of cognitive abilities (McGrew, 2009; Schneider & McGrew, 2012), and it contains three narrow abilities: induction, sequential (deductive) reasoning, and quantitative reasoning. Recently, Schneider and McGrew (2012) conducted a comprehensive review of the CHC theory and updated the model. The last model of CHC theory includes 18 broad cognitive abilities subsumed by more narrow abilities. Given the breadth of empirical support for the CHC intelligence structure, it offers one of the most valuable frameworks for designing and evaluating psychoeducational testing.

Figure 1
Cattell-Horn-Carroll (CHC) Model of Human Cognitive Abilities



Note. Gf = Fluid Reasoning, Gc = Comprehension Knowledge, Gwm = Short-Term Working Memory, Glr = Long-Term Memory, Gv = Visual Processing, I = Induction, RG = General Sequential (Deductive) Reasoning, RQ = Quantitative Reasoning. For simplicity, I only focus on fluid reasoning; thus, not all abilities are displayed.

Although fluid reasoning is only considered one of several abilities in the Cattell-Horn, Carroll, and CHC model, they all share a similar conclusion that fluid reasoning is the most important human cognition ability. Gustafsson (1984) found that the second-order factor of fluid reasoning is identical to the *g*-factor; thus, they should be considered the same factor. Cattell's investment theory can justify the observation that fluid reasoning is likely to correlate perfectly with *g*. The investment theory postulates that there is initially a single, general ability in the development of the individual related to the brain's maturation labeled as Gf. Through practice and experience, individuals develop abilities, and these developed abilities (i.e., Gc) are influenced by Gf and by struggle, motivation, and interest. The Gf develops into *g* because it influences the acquisition of knowledge and skills in different domains. For instance, most knowledge development comes from the inductive inference of partial knowledge found in different contexts (Landauer & Dumais, 1997). His further study (Kvist & Gustafsson, 2008) confirmed Cattell's investment theory suggesting the involvement of fluid reasoning in every aspect of the learning process.

Measuring Fluid Reasoning In A Multicultural Context

There are two traditions on conceptualizing fluid reasoning factors: process factors and content factors. These two conceptualizations are rooted on the same basis: a factorial analysis. Factor loadings are a component of the outcome of factor analysis, a data reduction technique to explain the correlations between variables with a reduced number of factors. Simply put, factor loading shows the extent to which the variable is related to a given factor (see Carroll (1993) and Wilhelm (2005) for further detail on the factor analytical technique). The process factors were classified based on the reasoning process, namely inductive (rule inference) and deductive (rule application). The content factors were classified based on the stimulus content of the test, namely figural, numeric/quantitative, and verbal. This conceptualization consequently affects how to measure fluid reasoning.

Carroll (1993) summary of factor analytics studies found three factors of fluid reasoning: inductive reasoning, sequential (deductive) reasoning, and quantitative reasoning. These factors also become the name used for narrow abilities of fluid reasoning in the CHC model (Schneider & McGrew, 2012). In addition, he suggested that inductive reasoning, which is mostly measured by the figural test, has the highest loading factor on fluid reasoning. A similar explanation was also found in Kan et al. (2011) reanalysis of the HCA (Human Cognitive Abilities project) data set. They found that the correlation between the first-order inductive reasoning factor and Gf was nearly one.

However, Wilhelm (2005) has a different argument on how to conceptualize fluid reasoning factors. He argued that introducing a distinction between inductive and deductive reasoning is unnecessary since they are perfectly correlated. It can be best interpreted as content factors rather than process factors, with verbal, figural, and numerical content factors determining fluid reasoning. In addition, he suggested that if researchers want to measure *g* with a single task, they should select a figural reasoning test since it has the highest loading on fluid intelligence. Although there are two perspectives on how fluid reasoning should be conceptualized, both agree that a figural test is the best practice to measure fluid reasoning.

Schneider and McGrew (2018) claimed that adding quantitative reasoning as a narrow ability muddies the categorization since it encompasses both inductive and deductive reasoning but is limited to quantitative information. On the other hand, inductive and deductive reasoning are process factors, whereas quantitative reasoning is a blend of process and content factors. They argued that figural, verbal, and numeric factors of fluid reasoning represent contaminants of visual-spatial processing, comprehension-knowledge, and quantitative knowledge. They then developed a model in which inductive reasoning is central to fluid reasoning and is largely unaffected by the kind of stimulus, but deductive reasoning is more dependent on content-specific processes. In addition, although there are two perspectives on how fluid reasoning should be understood (process vs content factors), both agree that a figural test is the best practice to measure fluid reasoning.

Several tests measuring fluid reasoning based on content factors have been developed for a specific purpose. For example, Berliner Intelligenzstruktur-Test der Deutschen Gesellschaft für Personalwesen (BIS-r-DGP); (Beauducel et al., 2001; Beauducel & Kersting, 2002), Cognitive Abilities Test Form 7 (CogAT 7), (Lohman, 2011) and the series and matrix test (Kyllonen et al., 2018). The benefit of measuring the three content factors of Gf is that it allows the test to align well with the reasoning demands of specific jobs or classroom activities, including considerable quantitative, verbal, and abstract reasoning (Beauducel et al., 2001). However, measuring content-based fluid reasoning also has several limitations, especially if it is carried out in a culturally different group. A cross-cultural study by Cockcroft et al. (2015) on the Wechsler Adult Intelligence Scales Third Edition (WAIS-III) found that the British group significantly outperformed the South African group on the knowledge-based verbal test. However, the groups did not differ significantly on the Matrix Reasoning subtest with minimal reliance on language, which appears to be the least culturally biased.

A culture-fair test is characterized as less specific to a culture (Cocodia, 2014). It implies that tests should accurately give scores that reflect the examinee's ability regardless of their cultural background. Greenfield (1998) suggested that a culture-fair test must include items requiring universal responses. It means that the subject response must be universally similar irrespective of ethnic group or culture. This culture-fair test is becoming more important in measuring fluid reasoning since fluid reasoning should be independent of acculturation. However, ensuring a verbal content test is universal for everyone is not easy. In the assessment, when the examinees come from various ethnic groups, the fluency of certain languages must be equal. When a person who is not fluent in a language is tested, the score may reflect the individual's language ability rather than their reasoning ability. Moreover, if one group is more familiar with the words than the other group, this test might be unfair for group comparison. Based on previous reasons, using a figural test seems more appropriate when the test is used for comparison in multicultural settings.

However, claiming that all figural tests are the culture-fair test cannot be true. Lakin and Gambrell (2012) faced a problem when they tried to develop a pictorial test to measure fluid reasoning. Culturally specific knowledge increases as they attempt to create more difficult items. A similar problem was faced by Schulze et al. (2005) when they developed picture-based analogy items for their study of adult intelligence. The test was based on cultural content (e.g., identification of Romanesque and Gothic architectural styles) that is not appropriate for operationalizing fluid reasoning measures. Culture fair tests try to minimize culture loading by not using pictures of familiar common objects. They only consist of simple elements - curves, lines, shapes, and only use simple concepts – left/right, up/down, smaller/larger, empty/full, and many/few. It must be noted that even if a culture-fair test does not remove the presence of culture on test items, tests are designed so they are common to various cultures.

In a worldwide context, several researchers even questioned whether a figural test is completely culture-fair (Carstensen et al., 2021; Ueda et al., 2017; Wicherts et al., 2010). Gonthier (2022) wrote an integrative review article discussing the cross-cultural differences in figural/visuospatial tests. He mentioned one example of the source of the unfairness of the tests. Most tests (e.g., series, matrices) require the examinee to process the stimulus from left to right. However, not all languages read left to right. Some languages, such as Hebrew and Arabic, read from right to left. These reading habits have a well-known impact on visual material exploration and could affect the examinee's performance. This argument is supported by the findings that individuals were better at recalling information if the spatial flow of presentation matched the dominant directionality of the individuals' culture (McCrink & Shaki, 2016). However, in a smaller context (e.g., within one country), usually, there are no such differences in reading habits. Thus, in this context, a figural test still guarantees culturally fair tests.

For several reasons previously mentioned, a figural test is the best practice to measure fluid reasoning regardless of prior experiences or the current ability for language. Measuring fluid reasoning using figural tests can reduce the cultural and language load of the tests. The use of fluid reasoning tests in the educational setting is, for example, to set up average intellectual levels of students. In addition, the fluid reasoning test is a good predictor of academic achievement (Cormier et al., 2016; Green et al., 2017; Solari et al., 2013), and it is useful for identifying gifted and talented students (Naglieri, 2016). In the Industrial and Organizational (I/O) context, psychological testing is mostly used in personnel selection to hire new employees. The empirical findings from many studies show a strong link between general cognitive ability and job performance (Schmidt, 2002). However, the selection process is costly and takes much time. Fluid reasoning measures can be the best candidate to estimate general cognitive ability due to the highest loading factor to general intelligence. Thus, measuring fluid reasoning is worthwhile under the assumption that novel problem-solving ability is commonly used in most jobs. In addition, using a figural reasoning test can be a better solution to answering fairness issues if the examinees come from various backgrounds.

Measuring Fluid Reasoning In Indonesia and Its Future Direction For Test Development

Indonesia is a multicultural country characterized by various religions, customs, and cultures, with over 1,300 unique native ethnic groups and over 580 local languages (Na'im & Syaputra, 2010). The official language of the nation is Indonesian, and practically all Indonesians speak it owing to its widespread use in school, business, and the media. However, 79.5 percent of the population interacts daily at home in their indigenous language (Na'im & Syaputra, 2010), which varies considerably amongst ethnic groups. Measuring the verbal content of Gf in Indonesia is difficult since it is conducted on a culturally distinct group.

Numerous challenges surround the culturally equitable measurement of Gf in Indonesia. First, Indonesia has a scarcity of cognitive testing. Until now, academics have relied on tests borrowed from Western nations, which are now considered obsolete. (e.g., the first version of Wechsler's test). Several Indonesian cognitive ability tests have been developed (e.g., the Indonesian Collective Intelligence Test (TIKI), Faxtor Cognitive Ability Test, and the AJT CogTest). Unfortunately, the technical report of the tests is inaccessible to the public, prohibiting researchers from analyzing the theoretical framework and its validity and reliability. Study on the existing tests is also limited. For example, I could not find any results when I searched in Garuda (platform for discovery of scholarly publications in Indonesia) using the keywords "*Tes Inteligensi Kolektif Indonesia*", "Faxtor Cognitive Ability Test", or "AJT CogTest". It indicates that the existing tests, although often used for practical purposes, are rarely evaluated or used for research purposes. It can be understood because access to existing tests is limited and expensive for Indonesian researchers.

Second, there is no culturally fair test measuring the process factor of fluid reasoning.

Specifically, no deductive reasoning tests are developed using non-verbal content. All deductive reasoning tests use verbal content, such as syllogism or arrangement subtest in Scholastic Aptitude Test ("*Tes Potensi Akademik*" in Indonesian). When a person who is not fluent in a language is tested using a verbal test, the score may reflect the individual's language ability rather than reasoning ability. Third, there is a need to develop a test with the flexibility and efficiency of testing processes that can be accessed freely.

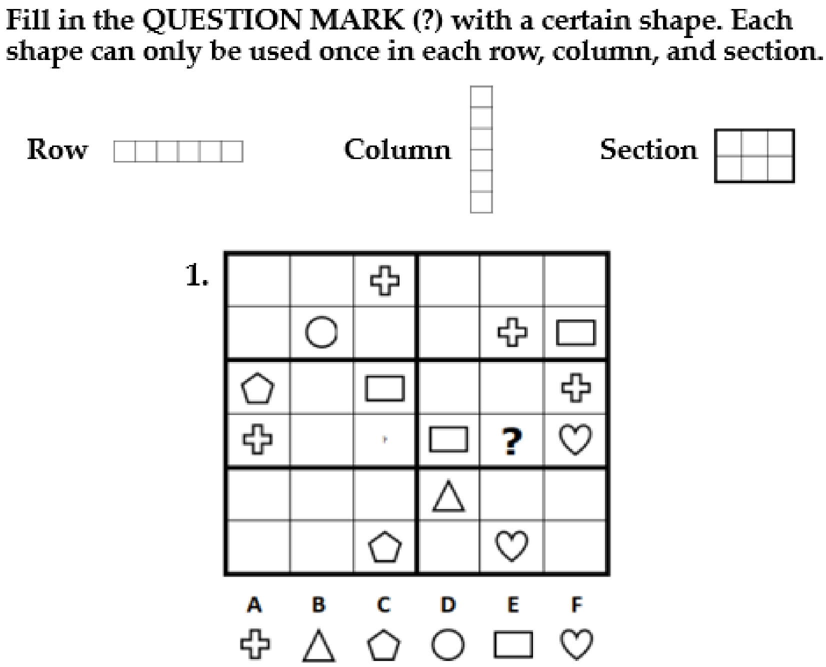
For Indonesian psychologists, RPM (Raven et al., 1988) is the most popular non-verbal test used to measure fluid reasoning. The examinees are asked to define the missing element that completes a pattern in each test item. Another popular test in Indonesia is CFIT (Cattell et al., 1973). The CFIT consists of three scales with non-verbal visual puzzles. In both scales 2 and 3, there are four subtests: series (completing a sequence of drawings), classification (choosing a picture that is different from others), matrix (completing a matrix of patterns), and conditions (choosing one picture that duplicates the conditions given). Although the first version of RPM and CFIT was developed over 70 years ago, they are still widely used today. RPM and CFIT as the solution for the more culturally fair test are overused in Indonesia, so the validity of the result is questionable. In addition, the RPM and CFIT only measure the process of inductive reasoning, whereas fluid reasoning could only be adequately represented if it was measured by at least two different narrow ability tests (Flanagan et al., 2013). Schneider and McGrew (2018) stated that no assessment of fluid reasoning is sufficient without a test of inductive reasoning, and the second should be a deductive reasoning test.

For several reasons, developing new tests benefits Indonesians. First, none of the non-verbal tests measures the comprehensive process of fluid reasoning (inductive and deductive reasoning). For instance, four subtests of CFIT only measure inductive reasoning, and RPM only measures Gf using a single task. The limits of any single test as an indicator of a latent factor have long been recognized, and researchers encourage expanding Gf measurement to different reasoning problem types (Hayes et al., 2015; Jaeggi et al., 2013; Schneider & McGrew, 2012; Stephens et al., 2020).

Second, most of the existing tests are commercial tests. The tests could only be used by paying a charge with a strict administration procedure by trained psychologists only. For clinical and industrial settings, copyright-protected commercial tests offer clear advantages. However, this nature of commercial tests is disliked by researchers, especially regarding the issues of cost and/or a lack of administrative flexibility. Conducting research becomes expensive, limiting the sample size. One example of a project such as the International Cognitive Ability Resource (ICAR) (Condon & Revelle, 2014) can be a solution for researchers to get the measurements they need.

Thus, developing multi-processes fluid reasoning tests is still a critical agenda today. I encourage researchers to develop new figural tests measuring two processes of fluid reasoning: inductive and deductive. Such a test will benefit the comprehensive measurement of fluid reasoning while ensuring culturally fair assessment. While almost all the existing non-verbal tests included inductive reasoning tests in their subtest, such as the CFIT subtests (e.g., matrices, series, classification, condition), they rarely included deductive reasoning. Deductive reasoning is typically measured by verbal content (e.g., syllogism test), which could be culturally biased. However, it is not impossible to measure deductive reasoning using a non-verbal test. For instance, Lee et al. (2008) suggested that individuals who solved the Sudoku puzzle depend on pure deduction. Thus, developing a Sudoku-like test could be a good candidate for a culturally fair measure of deductive reasoning. Figure 2 depicts one example of a sudoku-like item which could be a good candidate for measuring deductive reasoning.

Figure 2
Sample Item of Sudoku-Like Test



However, the challenge of developing new tests today requires a considerable investment of time and money and the researcher's expertise. Therefore, this idea has implications for several parties. Collaboration between researchers, universities, and test publishers is needed to achieve this goal. Universities and test publishers must provide funding for long-term research projects, while researchers must invest their time and expertise to develop the test. The final tests could be commercial, but an exemption should be made for research purposes. In the end, test publishers or universities can benefit by selling the tests, while researchers can benefit from the availability of the instrument for their research. Once the test is available, research on cognitive abilities in Indonesia, which is currently stagnant, can move forward.

Conclusion

Fluid reasoning is a central component of human abilities. Thus, measuring fluid reasoning is the best candidate to estimate general cognitive ability due to the highest loading factor to general intelligence. Although scholars from the psychometrics perspective mentioned that fluid reasoning and *g* are essentially identical (Gustafsson, 1984; Kvist & Gustafsson, 2008); (Blair, 2006) review from the developmental neuroscience perspective suggested that *Gf* and *g* should be distinguished. I argue that the interpretation of fluid reasoning measurement should be made based on the definition of fluid reasoning itself. In special cases when resources are limited and a single measure is needed (e.g., in personnel selection), using a fluid reasoning test can be the best candidate because of its central role in the structure of abilities. However, the interpretation should be made based on the assumption that novel problem-solving is commonly used in almost every job activity.

Recommendation

Another issue that should be resolved in the future is the lack of reasoning tests. Since verbal tests may be culturally specific and have a limitation in cross-cultural studies, developing a non-verbal test can be valuable. However, none of the non-verbal tests measures the comprehensive process of fluid reasoning. Ideally, a test is developed to represent different processes (inductive and deductive reasoning) and content (figural, verbal, numerical). However, in order to reduce cultural load, researchers are encouraged to develop figural tests with two process domains: inductive and deductive reasoning. Such a test, of course, will be helpful but still unavailable today for Indonesians. Collaboration between researchers, universities, and test publishers is needed to fulfill the gap. Future research could integrate test development with psychometrics and computer technology advancement. For instance, develop an automatic item generation, use computerized adaptive testing (CAT) for testing mode, and use multidimensional item response theory (MIRT) for the analysis.

Declarations

Acknowledgement

I am grateful to Kristof Kovacs for providing valuable feedback on earlier drafts of the manuscript.

Funding

The author received no funding support in preparing and publishing this article.

Conflict of Interest

There was no conflict of interest in the preparation and publication of this article

Author's Contribution

The present author is the sole author of this article.

Orcid ID

Hanif Akhtar https://orcid.org/0000-0002-1388-7347

References

Barkl, S., Porter, A., & Ginns, P. (2012). Cognitive training for children: Effects on inductive reasoning, deductive reasoning, and mathematics achievement in an australian school setting. *Psychology in the Schools*, *49*(9), 828–842. https://doi.org/10.1002/pits.21638

Beauducel, A., Brocke, B., & Liepmann, D. (2001). Perspectives on fluid and crystallized intelligence: Facets for verbal, numerical, and figural intelligence. *Personality and Individual Differences*, *30*(6), 977–994. https://doi.org/10.1016/s0191-8869(00)00087-8

Beauducel, A., & Kersting, M. (2002). Fluid and crystallized intelligence and the Berlin model of intelligence structure (BIS). *European Journal of Psychological Assessment*, *18*(2), 97–112. https://doi.org/10.1027//1015-5759.18.2.97

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences*, *29*(2), 109–125. https://doi.org/10.1017/s0140525x06009034

Blankson, A. N., & Blair, C. (2016). Cognition and classroom quality as predictors of math achievement in the kindergarten year. *Learning and Instruction*, *41*, 32–40. https://doi.org/10.1016/j.learninstruc.2015.09.004

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. In Cambridge University Press. Cambridge University Press.

Carstensen, A., Cao, A., Gao, S., & Frank, M. (2021). Investigating cross-cultural differences in reasoning, vision, and social cognition through replication.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22.

Cattell, R. B. (1983). *Intelligence: Its structure, growth, and action*. Elsevier science.

Cattell, R. B., Krug, S., & Barton, K. (1973). *Technical supplement for the culture fair intelligence tests, scales 2 and 3*. Institute for Personality; Ability Testing.

Cockcroft, K., Alloway, T., Copello, E., & Milligan, R. (2015). A cross-cultural comparison between South African and British students on the Wechsler Adult Intelligence Scales Third Edition (WAIS-III). *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00297

Cocodia, E. (2014). Cultural perceptions of human intelligence. *Journal of Intelligence*, *2*(4), 180–196. https://doi.org/10.3390/jintelligence2040180

Condon, D. M., & Revelle, W. (2014). The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52–64. https://doi.org/10.1016/j.intell.2014.01.004

Cormier, D. C., Bulut, O., McGrew, K. S., & Frison, J. (2016). The role of cattell–horn–carroll (CHC) cognitive abilities in predicting writing achievement during the school-age years. *Psychology in the Schools*, *53*(8), 787–803. https://doi.org/10.1002/pits.21945

Flanagan, D. P., Ortiz, S., & Alfonso, V. (2013). *Essentials of cross-battery assessment (3rd ed.)* John Wiley & Sons.

Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll Theory of cognitive abilities. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118660584.ese0431

Floyd, R. G., McGrew, K. S., & Evans, J. J. (2008). The relative contributions of the Cattell-Horn-Carroll cognitive abilities in explaining writing achievement during childhood and adolescence. *Psychology in the Schools*, *45*(2), 132–144. https://doi.org/10.1002/pits.20284

Frick, P. J., Barry, C., & Kamphaus, R. W. (2010). *Clinical assessment of child and adolescent personality and behavior*. Springer Science & Business Media.

Gonthier, C. (2022). Cross-cultural differences in visuo-spatial processing and the culture-fairness of visuo-spatial intelligence tests: an integrative review and a model for matrices tasks. *Cognitive Research: Principles and Implications*, *4*(7), 1–11. https://doi.org/https://doi.org/10.1186/s41235-021-00350-w

Green, C. T., Bunge, S. A., Chiongbian, V. B., Barrow, M., & Ferrer, E. (2017). Fluid reasoning predicts future mathematical performance among children and adolescents. *Journal of Experimental Child Psychology*, *157*, 125–143. https://doi.org/10.1016/j.jecp.2016.12.005

Greenfield, P. (1998). *The cultural evolution of IQ. In The Rising Curve: Long-Term Gains*. American Psychological Association.

Guerin, J. M., Sylvia, A. M., Yolton, K., & Mano, Q. R. (2020). The role of fluid reasoning in word recognition. *Journal of Research in Reading*, *43*(1), 19–40. https://doi.org/10.1111/1467-9817.12287

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*(3), 179–203. https://doi.org/10.1016/0160-2896(84)90008-4

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2015). Do we really become smarter when our fluid-intelligence test scores improve? *Intelligence*, *48*, 1–14. https://doi.org/10.1016/j.intell.2014.10.005

Horn, J. L., & Noll, J. (1997). *Human cognitive capabilities: Gf-gc theory*. Guilford.

Horn, J. L. (1967). Intelligence—why it grows, why it declines. *Society*, *5*(1), 23–31. https://doi.org/10.1007/bf03180091

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, *75*(3), 242–259. https://doi.org/10.1037/h0025662

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, *57*(5), 253–270. https://doi.org/10.1037/h0023816

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2013). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, *42*(3), 464–480. https://doi.org/10.3758/s13421-013-0364-z

Kan, K.-J., Kievit, R. A., Dolan, C., & van der Maas, H. (2011). On the interpretation of the (CHC) factor Gc. *Intelligence*, *39*(5), 292–302. https://doi.org/10.1016/j.intell.2011.05.003

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, *41*(7), 748–769. https://doi.org/10.1002/tea.20020

Kvist, A. V., & Gustafsson, J.-E. (2008). The relation between fluid intelligence and the general factor as a function of cultural background: A test of Cattell's Investment theory. *Intelligence*, *36*(5), 422–436. https://doi.org/10.1016/j.intell.2007.08.004

Kyllonen, P., Hartman, R., Sprenger, A., Weeks, J., Bertling, M., McGrew, K. S., Kriz, S., Bertling, J., Fife, J., & Stankov, L. (2018). General fluid/inductive reasoning battery for a high-ability population. *Behavior Research Methods*, *51*(2), 507–522. https://doi.org/10.3758/s13428-018-1098-4

Lakin, J. M., & Gambrell, J. L. (2012). Distinguishing verbal, quantitative, and figural facets of fluid intelligence in young students. *Intelligence*, *40*(6), 560–570. https://doi.org/10.1016/j.intell.2012.07.005

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295x.104.2.211

Lee, N. Y. L., Goodwin, G. P., & Johnson-Laird, P. N. (2008). The psychological puzzle of sudoku. *Thinking & Reasoning*, *14*(4), 342–364. https://doi.org/10.1080/13546780802236308

Lohman, D. F. (2011). *Cognitive abilities test, form 7*. Riverside Publishing.

McCrink, K., & Shaki, S. (2016). Culturally inconsistent spatial structure reduces learning. *Acta Psychologica*, *169*, 20–26. https://doi.org/10.1016/j.actpsy.2016.05.007

McGrew, K. S. (1997). *Analysis of the major intelligence batteries according to a proposed comprehensive gf-gc framework*. Guilford Press.

McGrew, K. S. (2009). 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

Motallebzadeh, K., & Yazdi, M. T. (2016). The relationship between (EFL) learners' reading comprehension ability and their fluid intelligence, crystallized intelligence, and processing speed (K.-w. Tong, Ed.). *Cogent Education*, *3*(1), 1228733. https://doi.org/10.1080/2331186x.2016.1228733

Naglieri, J. A. (2016). *Naglieri nonverbal ability test–third edition (NNAT3)*. Pearson.

Na'im, A., & Syaputra, H. (2010). Kewarganegaraan, suku bangsa, agama, dan bahasa sehari-hari penduduk indonesia [nationality, ethnicity, religion, and languages of indonesians]. *Statistics Indonesia*.

Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell-Horn-Carroll-based assessment. *Psychology in the Schools*, n/a–n/a. https://doi.org/10.1002/pits.20495

Raven, J., Court, J., & Raven, J. (1988). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. H. K. Lewis.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, *15*(1-2), 187–210. https://doi.org/10.1080/08959285.2002.9668091

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence.

Schneider, W. J., & McGrew, K. S. (2018). *The Cattell-Horn-Carroll theory of cognitive abilities*. Guilford.

Schulze, D., Beauducel, A., & Brocke, B. (2005). Semantically meaningful and abstract figural reasoning in the context of fluid and crystallized intelligence. *Intelligence*, *33*(2), 143–159. https://doi.org/10.1016/j.intell.2004.07.011

Solari, E. J., Aceves, T. C., Higareda, I., Richards-Tutor, C., Filippini, A. L., Gerber, M. M., & Leafstedt, J. (2013). Longitudinal prediction of 1st and 2nd grade english oral reading fluency in english language learners: Which early reading and language skills are better predictors? *Psychology in the Schools*, *51*(2), 126–142. https://doi.org/10.1002/pits.21743

Spearman, C. (1904). "General intelligence, " Objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201. https://doi.org/10.2307/1412107

Stephens, R. G., Dunn, J. C., Hayes, B. K., & Kalish, M. L. (2020). A test of two processes: The effect of training on deductive and inductive reasoning. *Cognition*, *199*, 104223. https://doi.org/10.1016/j.cognition.2020.104223

Ueda, Y., Chen, L., Kopecky, J., Cramer, E. S., Rensink, R. A., Meyer, D. E., Kitayama, S., & Saiki, J. (2017). Cultural differences in visual search for geometric figures. *Cognitive Science*, *42*(1), 286–310. https://doi.org/10.1111/cogs.12490

Wicherts, J. M., Dolan, C. V., Carlson, J. S., & van der Maas, H. L. (2010). Raven's test performance of sub-saharan africans: Average performance, psychometric properties, and the flynn effect. *Learning and Individual Differences*, *20*(3), 135–151. https://doi.org/10.1016/j.lindif.2009.12.001

Wilhelm, O. (2005). Measuring reasoning ability. In *Handbook of understanding and measuring intelligence* (pp. 373–392). SAGE Publications, Inc. https://doi.org/10.4135/9781452233529.n21