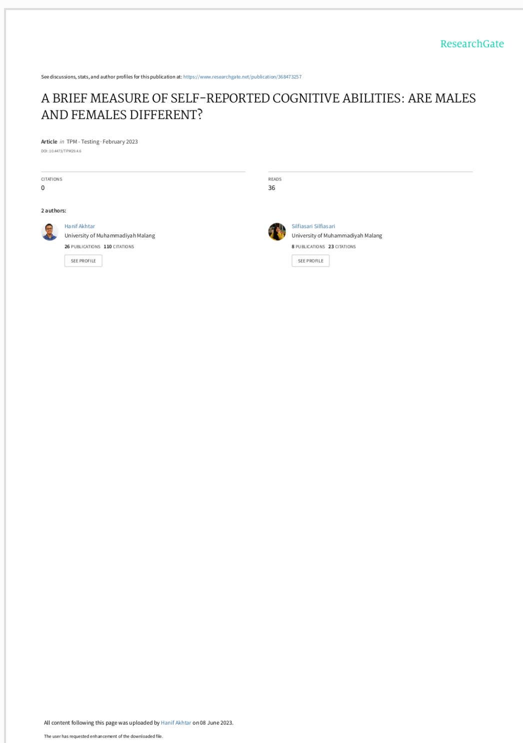
Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Silfiasari Silfiasari
Assignment title:	Publication Articles Juli-September 2023
Submission title:	A BRIEF MEASURE OF SELF-REPORTED COGNITIVE ABILITIES: ...
File name:	3._A_Brief_Measure_of_Self-Reported.pdf
File size:	637.46K
Page count:	20
Word count:	10,122
Character count:	54,864
Submission date:	19-Sep-2023 01:33PM (UTC+0700)
Submission ID:	2170418589



A BRIEF MEASURE OF SELF-REPORTED COGNITIVE ABILITIES: ARE MALES AND FEMALES DIFFERENT?

Article *in* TPM - Testing · February 2023

DOI: 10.4473/TPM29.4.6

2 authors:

Hanif Akhtar
University of Muhammadiyah Malang

Silfiasari Silfiasari
University of Muhammadiyah Malang

A BRIEF MEASURE OF SELF-REPORTED COGNITIVE ABILITIES: ARE MALES AND FEMALES DIFFERENT?

by Silfiasari Silfiasari

A BRIEF MEASURE OF SELF-REPORTED COGNITIVE ABILITIES: ARE MALES AND FEMALES DIFFERENT?

Article *in* TPM - Testing · February 2023

DOI: 10.4473/TPM29.4.6

2 authors:

Hanif Akhtar
University of Muhammadiyah Malang

Silfiasari Silfiasari
University of Muhammadiyah Malang

TPM

A BRIEF MEASURE OF SELF-REPORTED COGNITIVE ABILITIES: ARE MALES AND FEMALES DIFFERENT?

HANIF AKHTAR
ELTE EÖTVÖS LORÁND UNIVERSITY, BUDAPEST, HUNGARY
UNIVERSITY OF MUHAMMADIYAH MALANG, MALANG, INDONESIA

SILFIASARI
UNIVERSITY OF MUHAMMADIYAH MALANG, MALANG, INDONESIA

Recent advances in cognitive abilities theory provided by the Cattell-Horn-Carroll (CHC) model have made it essential to develop a new validated brief self-report measure of cognitive abilities. This paper outlines the development and initial validation of the Self-Report Cognitive Abilities Questionnaire (SRCAQ), a brief measure designed to assess cognitive functioning in the ability areas of fluid reasoning, short-term working memory, long term storage and retrieval, comprehension knowledge, processing speed, visual processing. Participants were 896 university students from Indonesia. The series of analyses using exploratory and confirmatory factor analysis confirmed the 6-factor structure of the SRCAQ. It had adequate factorial validity and demonstrated invariance across sex. This study also confirmed sex differences in self-reported cognitive abilities, with males tending to rate their fluid reasoning and visual processing higher than females. In the case of resources being limited, SRCAQ can be an alternative to estimate individuals' cognitive abilities. Limitations and suggestions for future research to expand the current study are presented.

Keywords: CHC theory; Self-report cognitive ability; Sex differences; Fluid reasoning; Visual processing.

Correspondence concerning this article should be addressed to Hanif Akhtar, Institute of Psychology, ELTE Eötvös Loránd University, Izabella Utca 46, 1064 Budapest, Hungary. Email: akhtar.hanif@ppk.elte.hu

For more than 40 years, several studies have focused on the issue of self-estimated or self-reported cognitive ability (intelligence). Self-estimated cognitive ability refers to a person's belief in their own abilities and plays a significant role in one's success in different areas of life (Chamorro-Premuzic & Furnham, 2006). This is based on the assumption that people are typically expected to be able to estimate their ability to a fair degree (Mirjalili et al., 2011). This topic is attractive for many reasons: to understand the individual's self-awareness and performance (Schlösser et al., 2013) and to understand individual differences that result in inaccuracies in assessing one's own cognitive ability (Paulhus et al., 1998). This topic has received numerous significant reviews (Szymanowicz & Furnham, 2013; von Stumm, 2014) and continues to attract many publications (Heck et al., 2018; Kaufman, 2012; Neto, 2019). It should be noted that most studies, including meta-analyses, find weak to moderate correlations between self-reported and performance-based cognitive abilities, with validity coefficients rarely reaching $r = .30$ (DeNisi & Shaw, 1977; Furnham, 2001; Paulhus, et al., 1998). Moreover, Herreen and Zajac (2018) found that personality was a stronger predictor of self-estimated cognitive abilities than actual cognitive performance, raising questions about the utility of this tool as a subjective measure of cognitive ability. Hence, the use of self-reported cognitive abilities cannot replace the measure of performance-based cognitive abilities and is not recommended in a competitive context.

The earliest studies on this topic looked at sex differences in cognitive ability estimation. Over the last four decades, scientists in differential psychology have reported relatively consistent sex differences in self-estimated cognitive ability. Specifically, males estimate their cognitive ability higher than females (for a review, see Furnham, 2001). This result has been replicated cross-culturally, according to a recent study comparing 12 nations from four continents (von Stumm et al., 2009). This research topic has received much attention and has been replicated in a lot of different countries, such as Austria (Stieger et al., 2010), Korea (Kang & Furnham, 2016), Brazil (Neto, 2019), Spain (Pérez et al., 2010), and Russia (Furnham & Shagabutdinova, 2012). A meta-analysis study was also conducted by Syzmanowicz and Furnham (2011) to examine the magnitude of sex differences in self-estimates of general, spatial, mathematical/logical, and verbal abilities. They found that in all but verbal ability, males estimated their ability significantly higher than females, with the mean effect size *d* for general intelligence being .37, for spatial .43, for mathematical .44, and for verbal .07.

Regarding the sex difference in self-reported intelligence (cognitive ability), Furnham and Grover (2020) summarized results from over 40 studies in this field and showed the following results: first, males of all ages tend to rate their intelligence more highly than females. These findings are consistent across cultures but vary depending on the type of intelligence being measured. For example, males rated their spatial and mathematical intelligence higher but their emotional intelligence lower than females. Second, people believe these sex differences occur across generations in one's family. Third, males estimate their intelligence to be higher than their original intelligence quotients (IQs), while females estimate their intelligence to be lower than their original IQs. Furnham (2001) suggested that people view intelligence as "male normative" because intelligence is more associated with mathematical, logical, and spatial abilities in which men are believed to be superior. Therefore, there is an assumption that men are smarter than women in society.

Although previous research reported relatively consistent sex differences in self-estimation of general intelligence, the difference between males and females in specific abilities is still unclear due to variation in the theoretical model of cognitive ability used. Most studies were conducted based on Gardner's (1993) multiple intelligence (MI) theory (Furnham, 2009; Furnham & Shagabutdinova, 2012; Furnham & Thomas, 2004; Kang & Furnham, 2016; Visser et al., 2008). According to the MI theory, intelligence is not an autonomous or solitary object. MI theory argues that every normal individual should develop, to some extent, some forms of intelligence (Gardner, 1993). Gardner (1999) recently defined intelligence as having eight forms: logical/ mathematical, verbal/linguistic, visual/spatial, bodily/kinesthetic, musical, intrapersonal, interpersonal, and natural intelligence. Individuals have different capacities and are not comparably identical in these abilities. This model of intelligence has been widely recognized in educational practice (Barrington, 2007). However, it has also been criticized due to a lack of empirical evidence and psychometrically valid measures (Waterhouse, 2010).

The Cattell-Horn-Carroll (CHC) model, in contrast to MI theory, offers researchers a cognitive ability model that is psychometrically validated (McGrew, 2009; Newton & McGrew, 2010; Schneider & McGrew, 2018). As a result, using this theoretical model to examine the accuracy of self-reported cognitive abilities might be beneficial. The last model of CHC theory includes 18 broad cognitive abilities subsumed by narrower abilities (Schneider & McGrew, 2018). Contemporary intelligence testing is often interpreted within the framework of CHC theory, and it has a huge impact on test interpretation (Kaufman, 2009). Under the CHC theory, the primary focus of interpretation is now widely recommended to be on normative strengths and weaknesses of scores reflecting broad and narrow abilities (Schneider & McGrew, 2012). The CHC model was used to classify subtests from all major intelligence test batteries in the late 1990s (Alfonso et al., 2005). Contemporary psychometric research has converged on the CHC model of cognitive abilities as the consensus working taxonomy of human cognitive abilities. Consequently, the current test developed to measure cognitive ability subscribes either explicitly or implicitly to CHC theory, including the self-reported cognitive ability. Contemporary cognitive ability tests (e.g., Woodcock-Johnson IV, Schrank & Wendling, 2018) measured at least seven broad abilities: fluid reasoning (Gf), short-term working memory (Gwm), long term storage and retrieval (Glr), comprehension knowledge (Gc), processing speed (Gs), visual processing (Gv), and auditory processing (Ga).

Currently, research in the measurement of self-reported cognitive ability based on the CHC theory framework has been conducted by several researchers. For example, Jacobs and Roodenburg (2014) developed a Self-Report Measure of Cognitive Abilities (SRMCA) to indicate the level of cognitive functioning in three CHC broad ability areas: Gf, Gc, and Gv. The SRMCA has been re-examined and revealed adequate reliability for each subscale of SRMCA ($\alpha > .7$) and convergent validity for the Gc domain but not for Gf or Gv (Herreen & Zajac, 2018). The other measures based on CHC theory were developed with more abilities involved (e.g., Fiorello et al., 2009; Waschbusch et al., 2000). Fiorello et al. (2009) developed a 37-item questionnaire measuring eight broad abilities of the CHC theory, while Waschbusch et al. (2000) developed a 50-item questionnaire measuring five broad abilities. However, both measures are not self-reported but others-reported. In addition, none of the existing studies investigated the measurement invariance across.

Two aspects of previous studies on this topic are worthy of attention and act as a motivation for the current study. First, currently, there is no brief measure that captures comprehensive self-reported cognitive abilities based on CHC theory. Existing measures, such as SRMCA (Jacobs & Roodenburg, 2014), only measure three broad abilities, while other measures (e.g., Fiorello et al., 2009; Waschbusch et al., 2000) are not self-report measures. A brief measure is essential for research since it allows for measuring multiple variables in a relatively short testing session and enables researchers to test more hypotheses (Gosling et al., 2003). Furthermore, many studies have compared the scores of self-reported cognitive abilities between males and females, but no study has proved sex invariance of the measure. It should be noted that no group comparisons are meaningful if the measure does not function equivalently across groups (Van De Schoot et al., 2015). Thus, before making any sex comparisons, researchers must investigate the sex invariance of the measure.

Second, most studies of self-reported cognitive abilities were conducted in Western countries, and only a small number of studies were conducted in Asia. Markus and Kitayama (1991) have suggested that people in Asia have different concepts of self-evaluation strategies and cognitive ability. Previous studies conducted in Japan and China have shown low-level self-estimates on all the types of intelligence rated, supporting the idea that how people estimate their level of ability is influenced by culture (Furnham & Fukumoto, 2008; Zhang & Gong, 2001). Currently, there is no study regarding sex differences in self-reported cognitive ability conducted in Indonesia. For years, Indonesia has been a male-dominated society, which may have influenced how Indonesian people perceive sex differences in cognitive abilities. Thus, replication of the study of sex differences in self-reported cognitive abilities in the Indonesian context is needed.

CURRENT STUDY

Given these considerations and previous studies on this topic, the goal of the current study is twofold. First, the current study aims to develop a Self-Report Cognitive Abilities Questionnaire (SRCAQ) for Indonesians based on the CHC theory framework. The initial 18 items were created to measure six broad abilities in the CHC theory framework (Gf, Gwm, Glr, Gv, Gc, and Gs). Those broad abilities were selected due to their central role in human cognition and are widely measured by contemporary cognitive ability tests. We used exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) to select the items, investigate and confirm the factor structure of the SRCAQ, evaluate its validity and reliability, and investigate the sex invariance of the questionnaire. Second, the current study aims to investigate sex differences in self-reported cognitive abilities based on the CHC theory framework. We hypothesize that our study would replicate previous studies on sex differences in self-reported cognitive ability. Namely, we hypothesize that males rate their cognitive ability higher than females, especially in fluid reasoning and visual processing.

STUDY 1
DEVELOPMENT OF A BRIEF MEASURE OF THE SELF-REPORT COGNITIVE ABILITIES QUESTIONNAIRE

Method

Participants

A total of 718 participants (345 males and 373 females) participated in this study. Recruitment occurred at the Faculty of Psychology, University of Muhammadiyah Malang, Indonesia; thus, all participants were undergraduate and master's students. Age ranged from 18 to 29 years (M = 19.4, SD = 1.13). The majority of participants were from Java (68.9%). Participants were recruited through the instructor's information in the class. The participants were given brief information related to the goal of the research. Fifty randomly selected participants would get a monetary reward of 50,000 Indonesian Rupiah each (3.5 United States Dollars). The participants completed the demographic questions and the 18 items of SRCAQ only if they were willing to participate in the research. The dataset was then randomly split into two parts: dataset A for EFA and dataset B for CFA. Therefore, each dataset contains 359 participants. This procedure was conducted to mitigate the problem of false-positive confirmation (Fokkema & Greiff, 2017).

Measure

All participants completed the 18-item SRCAQ. Items designed to measure Gf, Gwm, Glr, Gv, Gc, and Gs were created based on relevant CHC theory literature (Schneider & McGrew, 2012, 2018), existing self-report measures (e.g., Fiorello et al., 2009; Jacobs & Roodenburg, 2014; Waschbusch et al., 2000), and relevant subtests from popular cognitive abilities tests (e.g., Woodcock-Johnson IV, Schrank & Wendling, 2018). An expert in cognitive psychology reviewed a pool of 18 items for clarity, conciseness, and relevancy. For each item of the SRCAQ, participants rated how difficult or easy it was to perform specific cognitive tasks (for example, remember many things at once) compared to most people their age. Response options were as follows: 1 (*very difficult*); 2 (*difficult*); 3 (*medium*); 4 (*easy*); and 5 (*very easy*). The SRCAQ was pilot tested to ensure that the instructions, items, and response format were clear.

Data Analysis

Data were analyzed via exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). The analyses were run using the packages "psych," "lavaan," and "semTools" from the software R (Jorgensen et al., 2022; R Core Team, 2013; Revelle, 2022; Rosseel, 2012). The number of factors retained was based on parallel analysis and a theoretical basis. Parallel analyses are done on resampled data. The principal axis factoring extraction method combined with Promax rotation was used since permitting the factors to correlate was logically justified. To be retained, an item should exhibit a factor loading greater than .40 on the target factor or the factor loading on the target factor should be at least .15 greater than those on the other factors (Worthington & Whittaker, 2006). Cronbach's α was used to determine the subscales' reliability, with values higher than .70 being acceptable (Nunnally & Bernstein, 1994). In addition, we estimated the Spearman-Brown coefficient (r_{sb}) since it was recommended as the most appropriate reliability statistic for a 2-item scale (Eisinga et al., 2013).

The retained items were analyzed using CFA on a different dataset (dataset B). CFA was performed to check the overall model fit, convergent validity, discriminant validity, and construct reliability to confirm that the model accurately reflected the study constructs. The following fit indicators were calculated: χ^2/df, Tucker-Lewis index (TLI), comparative fit index (CFI), and the root-mean-square error of approximation (RMSEA). For a good model fit, χ^2/df should not exceed 3 (Kline, 2005), the TLI and CFI should exceed .95, and the RMSEA should not exceed .06 (Bagozzi & Yi, 1988; Hu & Bentler, 1999). Convergent and discriminant validity were evaluated by using criteria suggested by Fornell and Larcker (1981). Convergent validity was assessed by determining that the average variance extracted (AVE) for each construct was greater than the variance caused by measurement error for that construct (i.e., it should exceed .50). Discriminant validity can be assessed by comparing the AVE, maximum shared squared variance (MSV), and average shared squared variance (ASV). Discriminant validity is achieved when AVE exceeds MSV or ASV (Hair et al., 2018). Composite reliability values were used to determine reliability. Acceptable composite reliability is typically defined as .6 or above (Bagozzi & Yi, 1988). We used weighted least squares means and variance adjusted (WLSMV) estimation in our analyses as literature (e.g., Beauducel & Herzberg, 2006) suggested that this estimator is more appropriate for ordinal data.

A multigroup CFA was conducted on a total sample to investigate four levels of measurement invariance across sex: configural invariance, metric invariance, scalar invariance, and strict invariance. First, in configural invariance, all model parameters are freely estimated across sex. Second, in metric or weak invariance, the factor loadings are constrained to invariance across sex. Third, in scalar or strong invariance, the intercepts are set to be invariant in addition to the factor loadings. Scalar invariance makes comparing latent means across groups meaningful. Fourth, in strict invariance, the item residuals are constrained to be equal across groups. Once measurement invariance was met, a test of latent mean differences was conducted.

Results

Exploratory Factor Analysis (EFA)

The skewness and kurtosis of all items are between –1.00 and 1.00, indicating univariate normality. The average score for each item is 3.52. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy index was .891, and Bartlett's test of sphericity was significant ($p < .001$), suggesting that the sample was suitable for factor analysis. Parallel analysis was conducted, indicating a 4-factor solution. This result was counter-intuitive since the measure was designed to assess six latent constructs. We then ran the 4-factor and 6-factor models and used the fit statistic to determine the optimal number of factors to retain. Fit statistics of the 4-factor solution were TLI = .91; RMSEA [90% CI] = .54 [.042, .065], while fit statistics of the 6-factor solution was TLI = .96; RMSEA [90% CI] = .35 [.016, .050]. The model fit of the 6-factor solution has a lower RMSEA than the 4-factor solution (ΔRMSEA ≥ .015), so the 6-factor solution was then determined to represent the data best (Finch, 2020). The initial analysis of the 4-factor and 6-factor solutions to the 18-item SRCAQ is shown in Table C1 (Appendix C). As shown, there was an overlap between Gwm-Glr, Gf-Gv, and Gc-Gs items.

In the next stage, Items 3, 5, 7, 12, 15, and 18 were removed due to failure loads greater than 0.40 on the expected dimension. The EFA was rerun on the remaining 12 items for a final time, with six extracted factors explaining 55.85% of the variance. The correlations between the factors were moderate to high, ranging from r = .33 to r = .69 (Mean r = .51), indicating that the decision to use an oblique rotation was justified. The results of the final exploratory factor analysis are presented in Table 1. As shown in Table 1, items that were supposed to measure the same factor loaded on the same factor. All six scales failed to reach the accepted reliability α of .70. Cronbach's α for item groups for Gf, Gwm, Glr, Gv, Gc, and Gs α = .69, α = .65, α = .63, α = .66, α = .68, and α = .66, respectively. Spearman-Brown coefficient for item groups for Gf, Gwm, Glr, Gv, Gc, and Gs r_{sb} = .69, r_{sb} = .65, r_{sb} = .63, r_{sb} = .68, r_{sb} = .68, and r_{sb} = .66, respectively. It should be noted that the low reliability is due to the fact that the scales are very small (only two items for each scale).

TABLE 1
Loading factor derived from 12-item SRCAQ

Items	Factor					
	1	2	3	4	5	6
Come up with a solution to novel problems (SRCA1)	0.11	0.07	**0.62**	0.03	0.01	–0.05
Find the underlying principle of a problem (SRCA2)	–0.13	–0.13	**0.91**	–0.06	0.01	0.06
Remember many things at once (SRCA4)	0.01	**0.43**	0.10	–0.02	0.06	0.11
Remember the information despite many distractions (SRCA6)	–0.11	**1.03**	–0.13	–0.05	–0.02	–0.01
Remember lots of details about an event in the old photo (SRCA8)	**0.50**	0.14	0.00	0.06	–0.02	–0.03
Remember the names of my childhood friends (SRCA9)	**1.01**	–0.19	–0.08	–0.13	0.02	0.04
Determine if the furniture will fit in a room just by imagining it (SRCA10)	0.18	0.05	0.11	**0.49**	–0.03	–0.03
Imagine the shape of an object when it is translated, rotated, or reflected (SRCA11)	–0.15	–0.08	–0.09	**0.97**	0.05	0.05
Express a broad vocabulary (SRCA13)	0.04	–0.03	–0.06	0.05	**0.53**	0.19
Demonstrate the breadth of my general knowledge (SRCA14)	–0.02	0.02	0.06	0.01	**0.88**	–0.16
Read something quickly (SRCA16)	0.00	0.00	0.02	0.06	–0.13	**0.77**
Write fast (SRCA17)	0.03	0.03	0.03	–0.06	0.20	**0.55**

Confirmatory Factor Analysis (CFA)

CFA was conducted to compare which model of the three models proposed: the 6-factor model, the single-factor model, and the hierarchical factor model (see Appendix A, Figure A1) fit better with the data. We tested a single-factor model because the correlation between factors was moderate to high. In addition, a hierarchical factor (second-order) model was also tested since the measure heavily relied on the CHC model, which assumes that *g* underlies all broad factors. The summary of fit indicators of the models is presented in Table 2. The 6-factor model provided the best fit to the data, with the highest values of CFI and TLI and the lowest value of RMSEA, and SRMR, in comparison to a single-factor and second-order model. Since the comparison of three CFA models showed that the 6-factor model was the best fit, the following analyses were conducted on the 6-factor model.

TABLE 2
Model fit of three CFA models for the SRCAQ (N = 359)

Model	*df*	χ^2	CFI	TLI	RMSEA [90% CI]	SRMR
Single-factor model	54	231.96	.715	.652	.096 [.083; .109]	.087
Six-factor model	39	47.69	.986	.976	.025 [.000; .047]	.030
Second-order model	48	90.48	.932	.907	.050 [.034; .065]	.049

Note: CFA = confirmatory factor analysis; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; CI = confidence interval; SRMR = standardized root-mean-squared residual. All χ^2 tests were statistically significant at $p < .01$, all columns contain robust test statistics.

Table 3 shows all six measures' composite reliability, AVE, MSV, and ASV. The composite reliability ranges from .63 to .79, which is acceptable at .60 (Fornell & Larcker, 1981). The AVE ranges from .46 to .66. One factor (Glr) is below the recommended level of .50. According to Fornell and Larcker, if the AVE is less than .5, but the composite reliability is greater than .6, the construct's convergent validity is still adequate. Because the composite reliability of the six constructs is substantially above the required threshold, the internal reliability of the measurement items is adequate.

TABLE 3
Analysis of measurement accuracy

Factor	Indicator	λ	Z	CR	AVE	MSV	ASV	Factor					
								Gf	Gwm	Glr	Gv	Gc	Gs
Gf	SRCA1 SRCA2	.68 .73	10.79* 12.21*	.66	.50	.46	.28	**.71**					
Gwm	SRCA4 SRCA6	.81 .67	12.47* 9.02*	.71	.55	.27	.22	.45	**.74**				
Glr	SRCA8 SRCA9	.74 .61	11.93* 10.83*	.63	.46	.36	.24	.39	.52	**.68**			
Gv	SRCA10 SRCA11	.79 .83	15.61* 14.52*	.79	.66	.36	.22	.56	.35	.60	**.81**		
Gc	SRCA13 SRCA14	.70 .71	11.91* 11.11*	.66	.50	.48	.32	.68	.52	.47	.41	**.71**	
Gs	SRCA16 SRCA17	.82 .63	12.28* 9.35*	.69	.53	.48	.27	.53	.50	.45	.39	.69	**.73**

Note. λ = factor loading; CR = composite reliability; AVE = average variance extracted; MSV = maximum shared squared variance; ASV = average shared squared variance; square root of AVE = diagonal values in bold; off-diagonal value = correlation coefficients for each factor.
* $p < .001$

Discriminant validity was evaluated by comparing the AVE for each construct with MSV and ASV. The results of our analyses suggest that the indices of discriminant validity indicate good validity for all six factors (all AVE markedly higher than MSV and ASV). Overall, discriminant validity could be acceptable for this measuring model.

Invariance Analysis

A multigroup CFA was performed to investigate total samples' measurement invariance across sex. The invariance test examined four levels of measurement invariance: configural, metric, scalar, and strict. The configural invariance model (M1) is the first to be evaluated for these tests. We used Hu and Bentler's criteria of model fit indices to evaluate the configural model for each group: a CFI > .95, a TLI > .95, and an RMSEA < .06 (Hu & Bentler, 1999). Once the configural invariance model is accepted, the metric invariance model (M2) can be tested. This is investigated by constraining the factor loadings of items equally between the groups. If M2 does not differ from M1, metric invariance is inferred. Following this, the scalar invariance (M3) can be investigated. This is tested by applying the same factor loadings and intercept values to all groups. If M3 does not differ from M2, scalar invariance is inferred. Strict invariance (M4) is examined by constraining the factor loadings, intercepts, and residual variances to be equal across groups. If M4 does not differ from M3, strict invariance is inferred.

The configural model (M1) had adequate fit indices, $\chi^2 = 85.65$, $df = 78$, $p = .26$; CFI = .994; RMSEA [90% CI] = .017 [.000, .035], indicating the same factor structure in males and females. We assessed model fit differences and evaluated the following differences to indicate a lack of measurement invariance: ΔCFI ≥ .01; ΔRMSEA ≥ .015; ΔSRMR ≥ .010 (ΔCFI, ΔRMSEA, and ΔSRMR were interpreted in absolute values; Chen, 2007). The fit indices (e.g., chi-square, CFI, RMSEA, SRMR) of each model remained nearly unchanged in comparison to the fit indices of another model. Considering all of the results, the three models (M2 through M4) showed no significant reductions in model fits, indicating that the 6-factor structure reaches strict invariance between males and females (see Appendix B, Table B1). As a result, the 6-factor structure of the SRCAQ was shown to be sex invariant.

Latent Mean Comparisons

A comparison of latent mean differences across the male and female groups was conducted through multigroup CFA in the total sample. Latent mean values were set to zero in the female group and freely estimated for the male group. Because the latent means of the female group were fixed to zero, the latent means of the male group represented the mean differences between the two groups. The latent mean estimates were displayed in Table B2 (Appendix B), indicating mean values of Gf, Gv, and Gc in the male group were significantly different from zero ($p < .05$) but not in Gwm, Glr, and Gs. Males rated themselves as having higher Gf, Gv, and Gc than females. Cohen's (1988) *d* effect size index was used to transform the latent mean differences into a more familiar measure, following the procedure suggested by Hancock (2001). To determine if the variance values of the two factors are equivalent across sex, the variance values were fixed to be equal across the two groups. The model was then tested and compared from scalar invariance. The model had adequate fit indices, $\chi^2 = 125.88$, $df = 96$, $p = .022$; CFI = .978; RMSEA [90% CI] = .029 [.012, .043]; SRMR = .037. The difference between the CFI, RMSEA, and SRMR of the models examining the invariance of factor variances and scalar invariance was less than .01, thus indicating that factor variances were invariant across groups. As a result, the *d* values were calculated using the common standard deviations. The computed *d* values are shown in Table B2 (Appendix B).

STUDY 2
SEX DIFFERENCES IN SELF-REPORTED COGNITIVE ABILITIES

Method

Participants

A total of 178 participants (87 males, 91 females) participated in this study. Age ranged from 18 to 26 years (M = 19.4, SD = 1.37). Most participants were from Java (63.4%). Data collection was conducted in the Faculty of Psychology, University of Muhammadiyah Malang, and all participants were students in the "Scale Development" course. Participants took part in this study voluntarily as part of their learning activity in the course. Participants were given brief information about the study's goal and completed the 12 items of the SRCAQ.

Measure

All participants completed the 12-item SRCAQ. The description, development, and initial validation of this measure were mentioned in Study 1.

Data Analysis

The data for this study were analyzed using the multiple indicators and the multiple causes (MIMIC) model. A MIMIC model is an extension of CFA with covariates, usually demographic variables. By integrating these covariates into CFA under the latent variable framework, the associations between the latent variables and the demographic variables are simultaneously estimated with the factor loadings in the measurement model. In this study, the demographic variable that is hypothesized to affect the latent factors (self-reported cognitive abilities) is sex. First, the overall model was evaluated based on several fit statistics: χ^2, CFI, TLI, and RMSEA. For a good model fit, χ^2/df should not exceed 3 (Kline, 2005), the TLI and CFI should exceed .95, and the RMSEA should not exceed .06 (Bagozzi & Yi, 1988; Hu & Bentler, 1999). After the good model fit was achieved, we simultaneously tested the relationship between the latent factors of self-reported cognitive abilities and sex. The standardized effect size (Cohen's *d*) was calculated based on the procedure suggested by Hancock (2001). To compute standardize, we divide the unstandardized regression coefficient by the square root of the disturbance variance. Data analysis was performed using the "lavaan" package in R software (R Core Team, 2013; Rosseel, 2012).

Results

The WLSMV estimation converged to an acceptable solution for the MIMIC model. First, the overall model was evaluated based on the fit statistics described in Study 1. The MIMIC model, represented in Figure D1 (Appendix D), showed adequate fit indices, χ^2 = 46.38, df = 45, p =.41; CFI =.995; TLI =.992;

RMSEA [90% CI] =.013 [.000, .052]). Then, we investigated the associations between sex and the six latent factors. Sex significantly affected the Gf and Gv dimensions, showing that males reported higher mean scores for Gf and Gv than females. The standardized effect sizes (Cohen's *d*) were calculated based on the procedure suggested by Hancock (2001). To compute them, we divided the unstandardized regression coefficients by the square root of the disturbance variance. Table 4 depicts the associations between sex and the six latent factors in the MIMIC model.

TABLE 4
MIMIC model of self-repported cognitive abilities and sex

Path	Estimates	Standard error	*z*	*d*	*p*
Sex → Gf	0.24	0.11	2.05	0.40	.04
Sex → Gwm	0.01	0.14	0.08	0.01	.93
Sex → Glr	0.18	0.15	1.26	0.22	.21
Sex → Gv	0.32	0.13	2.34	0.41	.02
Sex → Gc	0.13	0.11	1.13	0.21	.26
Sex → Gs	–0.20	0.14	–1.38	–0.25	.17

Note. Females are coded as 0; thus, negative estimates indicate females have higher self-reported cognitive abilities than males. Gf = fluid reasoning; Gwm = short-term working memory; Glr = long term storage and retrieval; Gv = visual processing; Gc = comprehension knowledge; Gs = processing speed.

DISCUSSION

The first goal of this study was to develop SRCAQ, a new brief questionnaire measuring the self-reported cognitive abilities of Indonesians based on the CHC theory framework. To achieve this, we used EFA and CFA to select the items, investigate and confirm the factor structure of the SRCAQ, evaluate validity and reliability, and investigate the sex invariance of the questionnaire. EFA refined the SRCAQ from 18 to 12 items based on well-established classical test theory psychometric principles. It was anticipated that a 6-factor solution would describe the data, which was consistent with the theoretical basis proposed. The theorized 6-factor structure, however, did not emerge naturally. The initial analysis based on parallel analysis revealed a 4-factor solution with some overlap between items of Gf-Gv and Gc-Gs. The correlation between Gf and Gv is *r* = .58, while between Gc and Gs it is *r* = .69. This could be because Gf and Gv are typically highly correlated and deal with efficient and effective information processing (Chen et al., 2009), as do their corresponding SRCAQ items. At the same time, the high overlap between Gc and Gs items might be because the items of both abilities in this questionnaire are related to academic activities.

Further CFA analysis confirms the six factors (Gf, Gwm, Glr, Gv, Gc, and Gs) of self-reported cognitive abilities. With different datasets from the EFA, the CFA results revealed a satisfactory goodness-of-fit 6-factor model of SRCAQ with two items for each factor. By having only two items for each factor, this result is surprising since many researchers have recommended three to five items to represent each factor in factor analysis (MacCallum et al., 1999; Raubenheimer, 2004). However, this result is not uncommon since many studies (e.g., Islam, 2019; Muck et al., 2007; Nunes et al., 2018) found the same result in their analysis of the Ten-Item Personality Inventory (TIPI), a measure of Big Five personality traits with only two items each factor. Regarding sex invariance, we examined the configural invariance model, metric invariance model, scalar invariance model, and strict invariance model, successively. Finally, we confirmed that the 6-factor structure achieved strict sex invariance in Indonesian university students, which means that participants representing different sex interpret a given measure in a conceptually similar manner (Vandenberg & Lance, 2000).

Compared to other existing self-report measures based on CHC theory (Fiorello et al., 2009; Jacobs & Roodenburg, 2014; Waschbusch et al., 2000), this questionnaire has several benefits. First, it successfully confirms the 6-factor structure as well as the measurement invariance across sex. None of the previous studies proved sex invariance of the measure, even though this measure is frequently used for sex comparisons. Second, this questionnaire is shorter, with only 12 items measuring six broad abilities. Some benefits of using short measures are that they require less time, eliminate item redundancy, and reduce participant boredom (Gosling et al., 2003). However, compared to other instruments, this questionnaire has lower alpha reliability. It is almost impossible to get high alphas in instruments with only two items per factor unless the correlation between the items is very strong. However, a strong correlation between two items might be problematic because it indicates they are redundant. Researchers noted that calculating coefficient alpha on scales with a small number of items can be misleading (Kline, 2000).

Regarding the second goal of this study, we found significant differences across sex in self-reported Gf, Gc, and Gv in Study 1, and sex differences in Gf and Gv were replicated in Study 2, though with a small effect size. This finding supports a previous study conducted in Korea (Kang & Furnham, 2016) that found males have higher self-reported Gf than females. This finding is also consistent with a meta-analysis study (Syzmanowicz & Furnham, 2011), which synthesized data from predominantly Western nations (US, UK, and Europe) and found that males estimated their spatial and mathematical/logical abilities higher than females. Storek and Furnham (2013), in their research, confirm that intelligence is more closely related to masculinity. This study supports the idea that the sex effect on self-reported cognitive abilities, especially in Gf and Gv, is universal. Male samples in our study tend to evaluate themselves as superior in reasoning and spatial ability, which is more associated with general cognitive ability. Although our samples are not representative of the Indonesian population, they could partially imply that Indonesians still believe males outperform females in certain cognitive ability domains. It could explain why specific jobs that rely on reasoning and spatial ability (e.g., drivers, pilots, engineers) are dominated by males in Indonesia. Further study with a more representative sample is needed to confirm this finding.

Interestingly this finding is also in line with the study of sex differences in objectively measured cognitive abilities, which was tested by the Kaufman Assessment Battery for Children-Second Edition (KABC-II; Reynolds et al., 2008). They found that males consistently showed a significant mean advantage on the Gv factor at all ages. In a previous study, the self-reported Gv measure was correlated significantly and at a modest level with the performance Gv measure (r = .25) (Jacobs & Roodenburg, 2014). Although the correlation between self-reported and performance-based measures of cognitive ability is low, they share the same pattern in terms of sex differences. This finding raises the potential for further investigation of the relationship between objectively measured and self-reported cognitive abilities regarding sex differences. In particular, it is interesting to investigate what factors moderate individuals' evaluations of their abilities (e.g., personality, social desirability, and experiences with cognitive ability tests).

There were various limitations in this study. The main limitation of this study is the lack of any external validity measure. External measures could address the degree to which they correspond to any known cognitive or noncognitive constructs. However, we could not administer any external validation measures due to the limited resources and existing measures in Indonesia. The generalizability of the results is potentially limited due to the characteristics of the participants in this study. Our samples, although relatively evenly distributed in terms of sex, have a higher education level than the general population of Indonesians. Ability estimation may be affected by the homogeneity of the sample in terms of education. A previous study found some contradictory findings, especially in self-estimated general intelligence between college students and general samples. For example, Furnham and Shagabutdinova (2012), in their study on university students in Russia, revealed no sex difference in estimates of overall intelligence. Nevertheless, in their meta-analytical study, Syzmanowicz and Furnham (2011) found that males gave significantly higher self-estimates of general intelligence than did females ($d = 0.37$). University students may be more aware of their intelligence than the general population. However, these findings must be replicated before these speculative interpretations can be considered.

The other limitation of this study is the single administration of the measure to develop and validate the measure. Our study used two reliability coefficients: Cronbach's α and the Spearman-Brown coefficient. However, calculating alpha on scales with a small number of items can be misleading (Kline, 2000). In this case, providing another reliability index, such as test-retest reliability, was more appropriate. In general, more items lead to better construct representation, and the way to make measures more reliable is by increasing the number of items (Emons et al., 2007). For further research agenda, since broad abilities in CHC theory consist of many narrow abilities, we suggest developing a measure with more items that optimize content validity by including items that represent the narrow abilities.

Based on the results and limitations of our study, there are several suggestions for researchers as well as prospective users of this measure. To our knowledge, SRCAQ is the first measure of self-reported cognitive abilities available in Indonesia. SRCAQ could benefit from the advancement of self-reported cognitive abilities research in Indonesia. However, this brief measure should be limited to research purposes only, especially when the resources are limited. It is not intended for use in evaluation or self-assessment. In addition, this study provided initial validation of the SRCAQ. Further validation is needed, such as by correlating this measure with any existing objectively measured or self-reported measure of cognitive abilities.

Conclusion

In summary, a series of studies successfully confirmed the 6-factor structure of SRCAQ based on CHC theory. This measure has adequate factorial validity and demonstrates invariance across males and females. Thus, this brief questionnaire can be an alternative to estimate individuals' cognitive abilities when the resources are limited. Our study also confirmed previous studies regarding sex differences in self-report cognitive abilities. It showed that males tended to rate their Gf and Gv higher than females. Future studies are needed to replicate and extend the current findings, as the study reported here was the first to investigate the validity of the SRCAQ.

Acknowledgements

We would like to express our deep gratitude to Mr. Muhammad Salis Yuniardi for reviewing our initial items. We would also like to extend our thanks to Ms. Alifah Nabilah, Ms. Rizky Susanti, Ms. Istiqomah, and Ms. Noor Lusty, lecturer at the Faculty of Psychology, University of Muhammadiyah Malang, for their help in collecting the data.

REFERENCES

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 185-202). The Guilford Press.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74-94. https://doi.org/10.1007/BF02723327

Barrington, E. (2007). Teaching to student diversity in higher education: How Multiple Intelligence theory can help. *Teaching in Higher Education*, *9*(4), 421-434. https://doi.org/10.1080/1356251042000252363

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186-203. https://doi.org/10.1207/s15328007sem1302_2

Chamorro-Premuzic, T., & Furnham, A. (2006). Self-assessed intelligence and academic performance. *Educational Psychology*, *26*(6), 769-779. https://doi.org/10.1080/01443410500390921

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Chen, H. Y., Keith, T. Z., Chen, Y. H., & Chang, B. S. (2009). What does the WISC-IV measure? Validation of the scoring and CHC-based interpretative approaches. *Journal of Research in Education Sciences*, *54*(3), 85-108. https://doi.org/10.3966/2073753X2009095403004

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum.

DeNisi, A. S., & Shaw, J. B. (1977). Investigation of the uses of self-reports of abilities. *Journal of Applied Psychology*, *62*(5), 641-644. https://doi.org/10.1037/0021-9010.62.5.641

Eisinga, R., te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a 2-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637-642. https://doi.org/10.1007/s00038-012-0416-3

Emons, W. H. M., Sijtsma, K., & Meijer, R. R. (2007). On the consistency of individual classification using short scales. *Psychological Methods*, *12(*1), 105-120. https://doi.org/10.1037/1082-989X.12.1.105

Finch, W. H. (2020). Using fit statistic differences to determine the optimal number of factors to retain in an exploratory factor analysis. *Educational and Psychological Measurement*, *80*(2), 217-241. https://doi.org/10.1177/0013164419865769

Fiorello, C., Thurman, S., Zavertnik, J., Sher, R., & Coleman, S. (2009). A comparison of teachers' and school psychologists' perceptions of the importance of CHC abilities in the classroom. *Psychology in the Schools*, *46*(6), 489-500. https://doi.org/10.1002/pits.20392

Fokkema, M., & Greiff, S. (2017). How performing PCA and CFA on the same data equals trouble: Overfitting in the assessment of internal structure and some editorial thoughts on it [Editorial]. *European Journal of Psychological Assessment*, *33*(6), 399-402. https://doi.org/10.1027/1015-5759/a000460

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.2307/3151312

Furnham, A. (2001). Self-estimates of intelligence: Culture and gender difference in self and other estimates of both general (g) and multiple intelligences. *Personality and Individual Differences*, *31*(8), 1381-1405. https://doi.org/10.1016/S0191-8869(00)00232-4

Furnham, A. (2009). The validity of a new, self-report measure of multiple intelligence. *Current Psychology*, *28*(4), 225-239. https://doi.org/10.1007/s12144-009-9064-z

Furnham, A., & Fukumoto, S. (2008). Japanese parents' estimates of their own and their children's multiple intelligences: Cultural modesty and moderate differentiation. *Japanese Psychological Research*, *50*(2), 63-76. https://doi.org/10.1111/J.1468-5884.2008.00362.X

Furnham, A., & Grover, S. (2020). Correlates of self-estimated intelligence. *Journal of Intelligence*, *8*(1), 2-10. https://doi.org/10.3390/jintelligence8010006

Furnham, A., & Shagabutdinova, K. (2012). Sex differences in estimating multiple intelligences in self and others: A replication in Russia. *International Journal of Psychology*, *47*(6), 448-459. https://doi.org/10.1080/00207594.2012.658054

Furnham, A., & Thomas, C. (2004). Parents' gender and personality and estimates of their own and their children's intelligence. *Personality and Individual Differences*, *37*(5), 887-903. https://doi.org/10.1016/j.paid.2003.10.011

Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences* (2nd ed.). Fontana.

Gardner, H. (1999). *Intelligence reframed*. Basic Books.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504-528. https://doi.org/10.1016/S0092-6566(03)00046-1

Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2018). *Multivariate data analysis* (8th ed.). Cengage.

Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, *66*(3), 373-388. https://doi.org/10.1007/BF02294440
Heck, P. R., Simons, D. J., & Chabris, C. F. (2018). 65% of Americans believe they are above average in intelligence: Results of two nationally representative surveys. *PLoS ONE*, *13*(7), Article e0200103. https://doi.org/10.1371/journal.pone.0200103
Herreen, D., & Zajac, I. (2018). The reliability and validity of a Self-Report Measure of Cognitive Abilities in older adults: More personality than cognitive function. *Journal of Intelligence*, *6*(1), 1. https://doi.org/10.3390/jintelligence6010001
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118
Islam, N. (2019). The Big Five model of personality in Bangladesh: Examining the Ten-Item Personality Inventory. *Psihologija*, *52*(4), 395-412. https://doi.org/10.2298/PSI181221013I
Jacobs, K. E., & Roodenburg, J. (2014). The development and validation of the Self-Report Measure of Cognitive Abilities: A multitrait-multimethod study. *Intelligence*, *42*(1), 5-21. https://doi.org/10.1016/j.intell.2013.09.004
Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling. R package Version 0.5-6*. https://cran.r-project.org/package=semTools
Kang, W., & Furnham, A. (2016). Gender and personality differences in the self-estimated intelligence of Koreans. *Psychology*, *7*(8), 1043-1052. https://doi.org/10.4236/psych.2016.78105
Kaufman, A. S. (2009). *IQ testing 101*. Springer Publishing.
Kaufman, J. C. (2012). Self estimates of general, crystallized, and fluid intelligences in an ethnically diverse population. *Learning and Individual Differences*, *22*(1), 118-122. https://doi.org/10.1016/j.lindif.2011.10.001
Kline, P. (2000). *Handbook of psychological testing*. Routledge.
Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). The Guilford Press.
MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84-99. https://doi.org/10.1037/1082-989X.4.1.84
Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224-253. https://doi.org/10.1037/0033-295X.98.2.224
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
Mirjalili, R. S., Farahani, H. A., & Akbari, Z. (2011). Self-esteem as moderator of the relationship between self-estimated general intelligence and psychometric intelligence. *Procedia — Social and Behavioral Sciences*, *30*, 649-653. https://doi.org/10.1016/j.sbspro.2011.10.126
Muck, P. M., Hell, B., & Gosling, S. D. (2007). Construct validation of a short five-factor model instrument: A self-peer study on the German adaptation of the Ten-Item Personality Inventory (TIPI-G). *European Journal of Psychological Assessment*, *23*(3), 166-175. https://doi.org/10.1027/1015-5759.23.3.166
Neto, F. (2019). Sex differences in estimates of lay views about reversal motivational intelligences for self and others: A replication in Brazil. *Anales de Psicologia*, *35*(1), 68-74. https://doi.org/10.6018/analesps.35.1.302131
Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell-Horn-Carroll-based assessment. *Psychology in the Schools*, *47*(7), 621-634. https://doi.org/10.1002/pits.20495
Nunes, A., Limpo, T., Lima, C. F., & Castro, S. L. (2018). Short scales for the assessment of personality traits: Development and validation of the Portuguese Ten-Item Personality Inventory (TIPI). *Frontiers in Psychology*, *9*, Article 461. https://doi.org/10.3389/fpsyg.2018.00461
Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
Paulhus, D., Lysy, D. C., & Yik, M. S. M. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality*, *66*(4), 523-555. http://doi.org/10.1111/1467-6494.00023
Pérez, L. F., González, C., & Beltrán, J. A. (2010). Parental estimates of their own and their relatives' intelligence. A Spanish replication. *Learning and Individual Differences*, *6*(20), 669-676. https://doi.org/10.1016/J.LINDIF.2010.09.005
R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.r-project.org
Raubenheimer, J. (2004). An item selection procedure to maximise scale reliability and validity. *SA Journal of Industrial Psychology*, *30*(4), 59-64. https://doi.org/10.4102/sajip.v30i4.168
Revelle, W. (2022). *psych: Procedures for psychological, psychometric, and personality research. R package Version 2.2.5*. Northwestern University. https://cran.r-project.org/package=psych
Reynolds, M. R., Keith, T. Z., Ridley, K. P., & Patel, P. G. (2008). Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher-order MG-MACS and MIMIC models. *Intelligence*, *36*(3), 236-260. https://doi.org/10.1016/j.intell.2007.06.003

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02.
Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 597-599. https://doi.org/10.22237/jmasm/1257035100
Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the "signal extraction" counterexplanation for the Dunning-Kruger effect in self-evaluation of performance. *Journal of Economic Psychology*, *39*, 85100. https://doi.org/10.1016/j.joep.2013.07.004
Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 99-144). The Guilford Press.
Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & M. E. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73-63). The Guilford Press.
Schrank, F. A., & Wendling, B. J. (2018). The Woodcock-Johnson IV: Tests of cognitive abilities, tests of oral language, tests of achievement. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 383-451). The Guilford Press.
Stieger, S., Kastner, C. K., Voracek, M., von Stumm, S., Chamorro-Premuzic, T., & Furnham, A. (2010). Independent effects of personality and sex on selfestimated intelligence: Evidence from Austria. *Psychological Reports*, *107*(2), 553-563. https://doi.org/10.2466/04.07.09.PR0.107.5.553-563
Storek, J., & Furnham, A. (2013). Gender, *g*, gender identity concepts, and self-constructs as predictors of the self-estimated IQ. *Journal of Genetic Psychology*, *174*(6), 664-676. https://doi.org/10.1080/00221325.2013.772501
Syzmanowicz, A., & Furnham, A. (2011). Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta analyses. *Learning and Individual Differences*, *21*(5), 493-504. https://doi.org/10.1016/j.lindif.2011.07.001
Szymanowicz, A., & Furnham, A. (2013). Gender and gender role differences in self- and other-estimates of multiple intelligences. *Journal of Social Psychology*, *153*(4), 399-423. https://doi.org/10.1080/00224545.2012.754397
Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Measurement invariance [Editorial]. *Frontiers in Psychology*, *6*, Article 1064. https://doi.org/10.3389/FPSYG.2015.01064
Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4-69. https://doi.org/10.1177/109442810031002
Visser, B. A., Ashton, M. C., & Vernon, P. A. (2008). What makes you think you're so smart? *Journal of Individual Differences*, *29*(1), 35-44. https://doi.org/10.1027/1614-0001.29.1.35
von Stumm, S. (2014). Intelligence, gender, and assessment method affect the accuracy of self-estimated intelligence. *British Journal of Psychology*, *105*(2), 243-253. https://doi.org/10.1111/bjop.12031
von Stumm, S., Chamorro-Premuzic, T., & Furnham, A. (2009). Decomposing self-estimates of intelligence: Structure and sex differences across 12 nations. *British Journal of Psychology*, *100*(2), 429-442. https://doi.org/10.1348/000712608X357876
Waschbusch, D. A., Daleiden, E., & Drabman, R. S. (2000). Are parents accurate reporters of their child's cognitive abilities? *Journal of Psychopathology and Behavioral Assessment*, *22*(1), 61-77. https://doi.org/10.1023/A:1007576515188
Waterhouse, L. (2010). Multiple intelligences, the Mozart effect, and emotional intelligence: A critical review. *Educational Psychologist*, *41*(4), 207-225. https://doi.org/10.1207/S15326985EP4104_1
Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, *34*(6), 806-838. https://doi.org/10.1177/0011000006288127
Zhang, Y., & Gong, Y. (2001). Self-estimated intelligence and its related factors. *Chinese Journal of Clinical Psychology*, *9*, 193-195.

APPENDIX A

Confirmatory factor analysis (CFA) model comparison

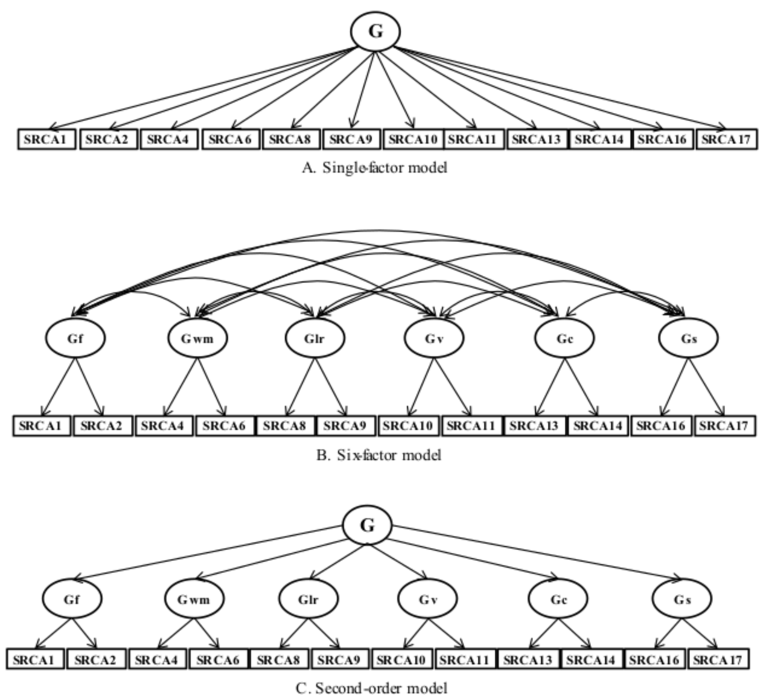


FIGURE A1
Three CFA models for the Self-Report Cognitive Abilities Questionnaire (SRCAQ)

APPENDIX B
Sex invariance testing

TABLE B1
Sex invariance testing of the SRCAQ

Model	Model fit				Model comparison		
	χ^2 (*df*)	CFI	RMSEA [90% CI]	SRMR	ΔCFI	ΔRMSEA	ΔSRMR
Configural (M1)	85.65 (78)	.994	.017 [.000, .035]	.027			
Metric (M2)	96.85 (84)	.990	.021 [.000, .037]	.030	.004	.004	.003
Scalar (M3)	105.80 (90)	.988	.022 [.000, .038]	.031	.002	.001	.001
Strict (M4)	123.51 (102)	.984	.024 [.000, .038]	.035	.004	.002	.004

Note. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval; SRMR = standardized root-mean-square residual.

TABLE B2
Results of latent mean difference analysis

Factor	Latent mean difference	*SD*	Cohen's *d*	*p*
Gf	0.18	.35	0.30	0.01
Gwm	0.05	.62	0.06	0.50
Glr	0.03	.61	0.03	0.72
Gv	0.15	.50	0.21	0.02
Gc	0.14	.38	0.23	0.01
Gs	–0.05	.47	–0.07	0.45

Note. Gf = fluid reasoning; Gwm = short-term working memory; Glr = long term storage and retrieval; Gv = visual processing; Gc = comprehension knowledge; Gs = processing speed.

APPENDIX C
Exploratory factor analysis

TABLE C1
Exploratory factor analysis results of 4- and 6-factor solutions

Items	4-factor solution				6-factor solution					
	1	2	3	4	1	2	3	4	5	6
Come up with a solution to novel problems (SRCA1)	0.21	–0.11	**0.60**	0.10	–0.09	**0.82**	0.01	–0.03	–0.11	0.18
Find the underlying principle of a problem (SRCA2)	0.28	–0.08	**0.57**	–0.17	–0.04	**0.84**	–0.01	–0.07	–0.03	–0.09
Solve complex problems with just a few hints (SRCA3)	**0.53**	–0.15	0.33	–0.02	**0.47**	0.35	–0.15	0.05	0.01	–0.09
Remember many things at once (SRCA4)	**0.65**	–0.03	0.01	0.13	**0.83**	–0.04	–0.07	0.01	0.01	–0.07
Recall the information I just received (SRCA5)	**0.45**	0.16	0.01	0.07	**0.33**	0.10	0.12	–0.10	0.16	0.06
Remember the information despite many distractions (SRCA6)	**0.55**	0.06	0.02	0.14	**0.61**	–0.06	–0.03	0.06	0.13	0.01
Remember the information I obtained a year ago (SRCA7)	**0.54**	0.03	–0.15	0.29	**0.69**	–0.09	0.12	–0.10	–0.10	0.08
Remember lots of details about an event in the old photo (SRCA8)	0.15	–0.08	0.02	**0.66**	0.09	–0.02	–0.01	0.01	0.00	**0.65**
Remember the names of my childhood friends (SRCA9)	0.11	–0.06	0.02	**0.54**	–0.09	0.01	–0.06	–0.11	0.12	**0.77**
Determine if the furniture will fit in a room just by imagining it (SRCA10)	–0.07	0.01	**0.48**	0.37	0.06	0.13	–0.03	**0.40**	–0.02	0.25
Imagine the shape of an object when it is translated, rotated, or reflected (SRCA11)	–0.18	0.21	**0.47**	0.17	–0.08	–0.12	0.02	**1.05**	0.07	–0.14
Remember the road I have been on (SRCA12)	–0.09	0.32	0.01	**0.33**	0.08	–0.05	**0.47**	0.09	–0.20	0.12
Express a broad vocabulary (SRCA13)	0.05	**0.82**	–0.16	0.01	–0.09	–0.07	**0.87**	–0.08	0.18	–0.10
Demonstrate the breadth of my general knowledge (SRCA14)	0.17	**0.52**	0.10	–0.04	0.07	0.14	**0.47**	0.03	0.15	–0.11
Understand verbal or written instructions (SRCA15)	0.11	**0.40**	0.17	0.05	–0.03	0.15	**0.32**	0.05	0.19	0.06
Read something quickly (SRCA16)	0.32	**0.31**	0.03	–0.10	0.00	–0.07	–0.02	0.06	**0.63**	0.07
Write fast (SRCA17)	**0.43**	0.41	–0.03	–0.15	0.00	–0.10	0.04	–0.03	**0.82**	0.07
Calculate simple math calculations quickly (SRCA18)	**0.26**	0.10	0.12	–0.06	0.10	0.13	0.00	0.02	**0.20**	–0.02

Note. In bold = highest loading factor; SRCA1-SRCA3 is intended to measure fluid reasoning (Gf); SRCA4-SRCA6 is intended to measure short-term working memory (Gwm); SRCA7-SRCA9 is intended to measure long term storage and retrieval (Glr); SRCA10-SRCA12 is intended to measure visual processing (Gv); SRCA13-SRCA15 is intended to measure comprehension knowledge (Gc); SRCA16-SRCA18 is intended to measure processing speed (Gs). Fit statistics of 4-factor solution, TLI = .91, RMSEA [90% CI] = .54 [.042, .065]; fit statistics of 6-factor solution, TLI = .96, RMSEA [90% CI] = .35 [.016, .050]

APPENDIX D

MIMIC model

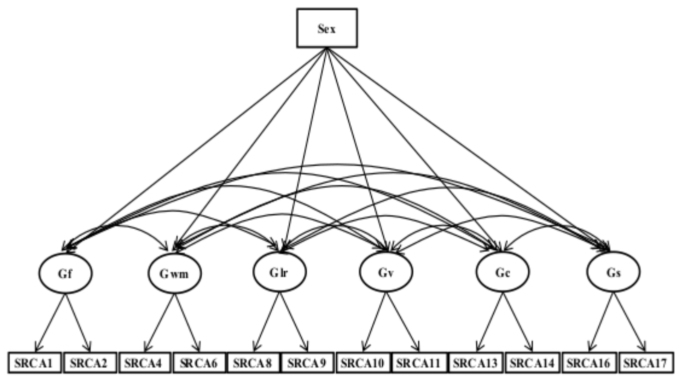


FIGURE D1
MIMIC model depicting the association between
the six factors of self-reported cognitive abilities and sex

A BRIEF MEASURE OF SELF-REPORTED COGNITIVE ABILITIES: ARE MALES AND FEMALES DIFFERENT?

ORIGINALITY REPORT

7%	6%	9%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	mdpi.com Internet Source	2%
2	Agata Syzmanowicz, Adrian Furnham. "Gender differences in self-estimates of general, mathematical, spatial and verbal intelligence: Four meta analyses", Learning and Individual Differences, 2011 Publication	1%
3	pdfs.semanticscholar.org Internet Source	1%
4	Gomez, R.. "Gender invariance of the five-factor model of personality among adolescents: A mean and covariance structure analysis approach", Personality and Individual Differences, 200609 Publication	1%
5	www2.mdpi.com Internet Source	1%
6	www.ncbi.nlm.nih.gov Internet Source	1%

7	Submitted to Eötvös Loránd University, Faculty of Education and Psychology Student Paper	1 %
8	digital.library.adelaide.edu.au Internet Source	1 %
9	publication.k-pin.org Internet Source	1 %

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		