Public Domain Intelligence Tests:

Psychometric properties of the Cog15 and ICAR16 cognitive ability scales

Dagný Kristjánsdóttir and Aya Zaiter

Department of Psychology, Lund University

PSYP01: Master's Thesis Work

Dr. Petri Kajonius

May 25th, 2023

Abstract

The current paper aims to explore the psychometric properties of two public domain cognitive ability scales, Cog15 and ICAR16, and investigate how well they each capture the g-factor in a Swedish sample ($N = 428$). The motivation for choosing these aims is that public domain, free, and easily accessible intelligence tests are needed for measuring the g-factor. Principal components analysis (PCA), confirmatory factor analysis (CFA) with maximum likelihood estimation, and reliability analyses returned results that indicated that ICAR16 is the better tool when measuring the g-factor, since it explained more of the variance (28.3%) and returned better reliability measures (Cronbach's $\alpha = .77$, McDonald's $\omega = .77$). We recommend omitting some of the items from the Cog15 and the ICAR16 scales. Future researchers should replicate the preliminary findings of this study on larger and more diverse samples to further understand the tests at hand, since Cog15 has yet to be researched as of now and ICAR16 is still under-researched.

Keywords: general intelligence, intelligence testing, public domain, CFA, PCA

History of Intelligence Tests

Despite the popularity and rigorous testing of intelligence tests (SB5, WISC-IV, WAIS-IV, etc.), they can be costly to take, ranging from $200 to $700, which limits their accessibility to most individuals (Bainbridge, 2020). This contributes to why the aim of this research is to assess the structural validity of the Cog15 and ICAR16 scales, two publicly available cognitive abilities tests. While the ICAR16 has been tested for its psychometric properties a few times, the Cog15 has not yet been validated as an effective measure of cognitive abilities (Kajonius, 2014).

Intelligence has been a topic of interest in research for several decades now (Colom et al., 2010). Differences in human mental ability or intelligence are distinct, significant, and arguable (Deary, 2001). The difficulty to define intelligence mainly stems from the challenge of defining the underlying constructs of intelligence (Colom et al., 2010). The most robust definition offered to date is that intelligence is hierarchical in nature and that there is a general factor of intelligence that is at the top of the hierarchy and encompasses more specific facets of intelligence, which is measured with IQ tests (Brody, 1999; Colom et al., 2010). The traditional measure of intelligence produces a single number that compares the results of a group of people who are the same age on a battery of subtests meant to measure various intellectual abilities (Brody, 1999). Over the past century, psychologists have developed and validated a variety of standardized tests aiming to measure intelligence (Colom et al., 2010). Some of the first and most widely used intelligence tests were the Stanford-Binet Intelligence Scale and the Wechsler Intelligence Scales (Lassiter et al., 2001; Silverman et al., 2010).

According to a review conducted by Bain and Allin (2005), the Stanford-Binet Intelligence Scale (SB5) is a test to measure cognitive abilities and intelligence for people across different age groups. It contains five factors across the verbal and non-verbal domains of the test, which are fluid reasoning, knowledge, quantitative reasoning, visual-spatial processing, and working memory (Bain & Allin, 2005). Evidence has particularly shown that the SB5 is a strong measure of general intelligence, namely in children and adolescents using a single factor, therefore, it is recommended that the scores are interpreted as overall IQ scores (Canivez, 2008). The SB5 can also be used as a diagnostic tool for potential cognitive or developmental delays in young children (Bain & Allin, 2005). The SB5 takes generally less time (15 – 75 minutes) to administer compared to previous versions of the Stanford-Binet intelligence test, and it is considerably easier to administer to different age groups (Bain & Allin, 2005).

Additionally, the Wechsler Intelligence Scale for Children (WISC) was published as an extension of the original Wechsler-Bellevue Intelligence Scale (Seashore et al., 1949). However, it is a distinct and independently standardized intelligence measure (Seashore et al., 1949). The original WISC contains 12 tests which are divided into two groups; Verbal and Performance (Seashore et al., 1949). The latest revision of the scale, WISC-IV contains 10 core subtests that are organized into the following four indexes; verbal comprehension, perceptual reasoning, working memory, and processing speed (Kaufman et al., 2006). The WISC-IV places less emphasis on timed tasks, compared to other intelligence tests, the only sections of the WISC-IV which have time limits are those that aim to measure speed (Kaufman et al., 2006). The WISC-IV has robust psychometric properties, with outstanding construct validity; it provides good measurements of both theory-based and research-based constructs of fluid intelligence (Gf) as well as working memory (Kaufman et al., 2006).

Finally, in an article done by Lassiter et al. (2001), the Wechsler Adult Intelligence Scale (WAIS) was named one of the most commonly used intelligence tests administered to adults. The WAIS-IV includes 15 subtests in total; 10 core subtests and 5 supplemental subtests (Gignac & Watkins, 2013). These subtests were designed to measure four inter-correlated factors, including Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed, and considering they are inter-correlated, they form a total scale score (FSIQ) (Gignac & Watkins, 2013). The WAIS also presents strong psychometric properties, including high internal consistency reliability (Lassiter et al., 2001).

The General Factor (g-Factor)

The previously mentioned intelligence tests aim to measure general global intelligence (Bain & Allin, 2005; Canivez, 2008; Gignac & Watkins, 2013; Kaufman et al., 2006; Lassiter et al., 2001; Seashore et al., 1949). Intelligence is composed of over sixty cognitive abilities, but in the early 20th century, the English psychologist, Charles Spearman was the first to introduce the general factor (g-factor or g), which combines all cognitive ability factors into one factor (Carroll, 1993; Spearman, 1904). Spearman conducted experiments to determine if certain cognitive abilities are inter-correlated, using a sample of school children from a small village (Carroll, 1993). Tests of verbal and spatial skills as well as other aptitudes were conducted (Spearman, 1904). Factor analysis further confirmed that the general factor of intelligence, the g-factor, could account for the empirically observed positive correlations between various cognitive tests (Kovacs & Conway, 2016). Spearman (1904) recognized that there was a chance the general factor might not be able to fully account for the variation in mental ability tests. His method of removing variables from correlation matrices that produce nonvanishing tetrad differences led to the discovery of tests that measure abilities beyond general factors (Spearman, 1904).

Spearman’s g-factor has faced some criticism, there among from the American psychologist Louis Leon Thurstone, who put forth a rival model made up of a number of unrelated group factors that represented a set of "Primary Mental Abilities" (Thurstone, 1938; Kovacs & Conway, 2016). Thurstone’s model contains seven different factors; word fluency, verbal comprehension, spatial visualization, number facility, associative memory, reasoning, and perceptual speed (Thurstone, 1938). Though Thurstone challenged Spearman, his original model was also put to the test, the notion of orthogonal factors proved to be unsupportable, and their correlations needed to be explained by a higher order general factor (Carroll, 1993).

Spearman's work was the first of its kind, and even though it was not flawless, including weak data and some calculation errors of the correlations, its importance to the research field of human intelligence and cognitive abilities cannot be disputed (Carroll, 1993). The g-factor is more highly correlated with various measures of learning, performance, and achievement than any other factor or combination of factors derived from the factor analysis of a given set of tests (Jensen, 1992). The g-factor normally accounts for about 40% or more of the variance in cognitive ability tests that are measured in large samples (Deary et al., 2010). The stability of the g-factor is evident across various factor analytic algorithms, test batteries, and populations (Bock et al., 2000). All cognitive tests, despite the varying levels of informational content, display some degree of the g-factor (Bock et al., 2000). This indicates that the g-factor cannot be attributed to test content or psychological factors, but rather is an inherent property of the brain (Bock et al., 2000). The g-factor has also been found to correlate with non-psychometric variables to a certain extent (Bock et al., 2000). The g-factor was found to be correlated with genetic heritability of intelligence test performance, brain size, and reaction time (Bock et al., 2000). The g-factor also appeared to be linked to most cognitive functions that establish individual differences and affect speed of information processing (Bock et al., 2000). De La Fuente et al. (2021) have further confirmed that a genetic g-factor exists. Twin studies that have used multivariate methods to examine genetic associations within cognitive test score variance suggest that there is strong heritability for the g-factor (De La Fuente et al., 2021). The results indicate that some cognitive traits are positively correlated due to having overlapped genetic makeups (De La Fuente et al., 2021).

New Public Domain Intelligence Tests: Cog15 and ICAR16

As previously mentioned, public domain tests are rare, as most intelligence tests are usually costly to administer (Bainbridge, 2020). A newly developed cognitive ability test is the Cog15, a freely accessible, public domain intelligence test created using classical IQ problems (Kajonius, 2014). It is a 15-item, performance-based, cognitive ability test, and the questions used have been adjusted from public material to measure four factors; numerical series, pattern reasoning, spatial ability and verbal skills (Kajonius, 2014). The minimum score is zero while the maximum score is 15 (Kajonius, 2014).

Another one of the few publicly available intelligence tests is the International Cognitive Ability Resource (ICAR16), a scale developed to measure general cognitive abilities, and is administered with no time limit (Condon & Revelle, 2014). The ICAR16, also named the ICAR Sample Test, is a shortened subset of the original ICAR60 (Condon & Revelle, 2014). It includes four item types, letter and number series, matrix reasoning, verbal reasoning, and 3D rotation (Condon & Revelle, 2014). According to Condon and Revelle (2014), the letter and number series items entail a sequence of numbers or letters and require the participants to identify the next item in the sequence based on six proposed options. The matrix reasoning items include a grid of nine geometric shapes with one of the shapes missing; the participants would need to identify the correct shape that completes the grid from a choice of six different shapes (Condon & Revelle, 2014). The verbal reasoning items entail general knowledge questions that aim to assess logic and vocabulary skills (Condon & Revelle, 2014). The 3D rotation items display different cube images and ask participants to choose which option is a potential rotation of the stimulus at hand (Condon & Revelle, 2014).

A study conducted by Young and Keith (2020) examined the construct validity of the ICAR16 and its relation to overall intelligence and cognitive abilities, by administering the test to a sample of 97 university students. Confirmatory factor analyses were conducted on the data obtained from the sample and results indicated that the ICAR16 was the most highly correlated with fluid reasoning and visual-spatial processing and was the least correlated with working memory and processing speed (Young & Keith, 2020). From the subscales included in the ICAR16, the letter-number series subscale best measured fluid reasoning, whereas the other three subscales were better measures of visual-spatial processing (Young & Keith, 2020). However, given the modest sample size, Keith and Young (2020) highlighted that the conclusions reached within this study were preliminary and could be biased, and recommended replication of the analyses on larger and more diverse samples. Another previous study of the ICAR16 was conducted on a Danish sample (Kirkegaard & Nordbjerg, 2015). A Danish translation of the original scale was used to administer the questionnaire online ($N = 72$). This study revealed that one general factor is present in the translated version of the ICAR16 (Kirkegaard & Nordbjerg, 2015). Additionally, the Cog15 scale has a moderate correlation ($r = .61$) with the ICAR16 scale (Kajonius, 2014).

Aim and Significance

There is a need for public domain, free to use and easily accessible intelligence tests to measure the g-factor. Our study aims to assess the effectiveness of Cog15 and ICAR16 as measures of cognitive abilities. The Cog15 scale is yet to be validated for its psychometric properties. On the other hand, the ICAR16 has been previously tested, but most of the structural validity analyses were conducted on smaller sample sizes (Kirkegaard & Nordbjerg, 2015; Young & Keith, 2020), and this study marks the first investigation of the ICAR16 scale in a Swedish population. Hence, this paper aims to fill a gap in the literature by examining the psychometric properties of two open-access intelligence tests and comparing them to determine which test is more adept at measuring the g-factor.

Methods

Participants

The participants were individuals from Sweden interested in completing their cognitive ability profile. Since the research was done in Sweden, Swedes were the most accessible sample for data collection. The study used a cross-sectional research design, mainly because it is cost-effective, with a lower dropout rate since people are only asked to participate one time (Taris et al., 2021). All data was collected ahead of time and is pre-existing at the time of this research being conducted. Participants were recruited through an open website link that was running for a couple of years. Participants who did not complete 90% (i.e., 31 out of 34 items) or more of the remaining sections of the survey were excluded from the study.

The final sample size was $N = 428$. Majority of the participants identified as female (66%), while the rest identified as male (33%). There were three participants who did not disclose their gender identity. The age of the participants ranged from 19 - 61 years old ($M = 27.9$, $SD = 8.6$), with 14 participants who did not disclose their ages.

Materials

The scale used in the data collection process was a combination of the Cog15 and the ICAR16 (see Appendix A). The survey included demographic questions, such as gender and age, 15 questions from the Cog15, and the 16 ICAR Sample Test items. The Cog15 has acceptable reliability, with Cronbach's $\alpha = .71$ (Cronbach, 1951; Kajonius, 2014). The Cog15 does not have any previously reported omega coefficient. The ICAR16 has good reliability, with Cronbach's $\alpha = .81$ and McDonald's $\omega = .83$ (Condon & Revellse, 2014; Cronbach, 1951; McDonald, 1999). The items included in the scoring process were three rationality questions (3 points), the Cog15 items (15 points), and the ICAR16 items (16 points), meaning the full score that could be achieved on this test was 34 points, and the minimum was 0. More specifically, the Cog15 scores can range from 0 to 15 points, whereas the ICAR16 scores can range from 0 to 16 points.

Analysis

The dataset obtained from the test which was distributed was analyzed in the statistical software Jamovi 2.3.21 (The jamovi project, 2022), IBM SPSS 29.0 (IBM Corp., 2022), and AMOS 29.0 (Arbuckle, 2022).

Categorical Variables

All answers obtained from the Cog15 and ICAR16 were recoded into 0 = *incorrect answer*, and 1 = *correct answer*, making the data collected binary. A descriptive analysis was performed on all the items of Cog15 and ICAR16, examining their mean value (see Table 1). Furthermore, histograms of the total score for each scale were computed (see Figures B1 and B2) to better visualize the distribution of the scores across the sample. Finally, correlation tables for the scales were also produced to see how the different variables inter-correlate (see Tables B1 and B2).

Principal Components Analysis (PCA)

For both Cog15 and ICAR16, a principal components analysis (PCA) was conducted on the first half of the sample (nPCA = 213), to assess which scale is better at capturing the g-factor. The dataset was split in half as a way of avoiding redundant analyses. PCA is one of the most commonly used multivariate statistical tools, across several different scientific disciplines (Abdi & Williams, 2010). PCA aims to extract important information from a dataset, reduce the size of the dataset, provide simple descriptions of the data, as well as analyze the structure of the variables (Abdi & Williams, 2010). PCA typically analyzes data described by a group of dependent variables, which are usually inter-correlated (Abdi & Williams, 2010). It attempts to extract the most relevant information in the form of new orthogonal variables, also known as principal components (Abdi & Williams, 2010).

However, to check that the datasets are suitable for factor analysis, some assumptions need to be checked; included in the PCA will be two measures that ensure a factor analysis is feasible. The adequacy of the sample is measured by Kaiser-Meyer-Olkin's (KMO) measure, which indicates whether factor analysis is suitable for the sample at hand (Kaiser, 1974). If the value of KMO is larger than 0.5, then the sample is considered adequate to go through factor analysis (Field, 2000). Additionally, Bartlett's test of sphericity measures the strength of the relationship between the variables of a scale (Bartlett, 1954). For the Bartlett test to be significant, its p-value needs to be less than .05 and that suggests that the data is approximately normally distributed, and therefore, suitable for factor analysis (Field, 2000).

Confirmatory Factor Analysis (CFA)

Following the PCAs, confirmatory factor analyses (CFA) were performed on the other half of the sample (nCFA = 215) for both Cog15 and ICAR16. The CFA was performed to confirm that the factor drawn from the PCA is valid. CFA is one of the most frequently used statistical procedures in applied research (Brown, 2015). CFA is a structural equation modeling (SEM) that is used to analyze relationships between observed measures or indicators and latent variables or factors, based on a theoretical background (Brown, 2015; DiStefano & Hess, 2005). For all PCAs and CFAs performed, the cut-off for the standardized factor loadings was set to .32, as recommended by Tabachnick & Fidell (2007), within samples larger than 300. Factor loadings are defined as the extent to which a variable or item is contributing to a factor; the larger the loadings are, the more the variables contribute to said factor (Yong & Pearce, 2013).

Additionally, the maximum likelihood method (ML) was used for the CFA. Based on a multivariate normal distribution, maximum likelihood estimates factor loadings (Tucker & Lewis, 1973). It is a method that maximizes the likelihood function to find the best solution (Tucker & Lewis, 1973). Maximum likelihood estimation is the best method for factor extraction when the data is relatively normally distributed (Fabrigar et al., 1999). It allows the calculation of a wide range of indexes of the model's goodness of fit and allows the computation of confidence intervals based on statistical significance tests of factor loadings and correlations among factors (Fabrigar et al., 1999). Moreover, oblique rotations were used as they are able to accurately model both correlated and uncorrelated factors, unlike orthogonal rotations, which are less effective in dealing with correlated factors (Osborne, 2015). Furthermore, it is more suitable to use oblique rotations, considering the results would be unaffected if the factors had a correlation of 0, it would simply produce a similar outcome to that when using orthogonal rotations (Osborne, 2015).

Concerns with Categorical Variables

One concern which could arise when using factor analysis for binary data is the fact that the assumptions needed to conduct a factor analysis could be violated by the nature of the data (Silvia et al., 2012). When factor analysis is conducted on binary datasets, it could be likened to Item Response Theory (IRT; Silvia et al., 2012). One recommendation to ensure that the binary nature of the data does not compromise the results obtained from the factor analysis was to use maximum likelihood estimation and taking standard errors into account, and using oblique rotations when conducting the analyses (Silvia et al., 2012). Another suggestion made by Silvia et al. (2012) was to convert the categorical variables into numerical values, which was done in the current study, by recoding all answers obtained on the scales into numerical data points.

Reliability Measures

Moving forward, testing the psychometric properties of the scales we used two measures, Cronbach’s alpha (α) and McDonald’s omega (ω) (see Tables 2 and 3) (Cronbach, 1951; McDonald, 1999). Cronbach’s alpha is an internal consistency measure for tests and scales and ranges between 0 and 1 (Cronbach, 1951). Additionally, according to Tavakol and Dennick (2011), it measures the extent to which items in a test measure the same concept. Acceptable values of alpha range between .70 and .95 (Tavakol & Dennick, 2011). As previously stated, coefficient omega (*ω*; McDonald, 1999) was calculated to test the psychometric properties of the scales. In terms of reliability, coefficient omega is a factor-analytic model that measures the overall variance in the data (McDonald, 1999; Rodriguez et al., 2016). Values exceeding .75 classify as acceptable omega coefficients (Reise et al., 2013).

Ethics

The study was conducted in accordance with the national regulations concerning research ethics. Considering there was no foreseeable risk with participation, an ethical review was not required. All participants were informed of the purpose of the study. Participants were also informed that participation in the study was completely voluntary, it was made clear that they could drop out of the study at any time without providing any explanation.

Results

The results of the descriptive analysis performed on Cog15 and ICAR16 are shown in Table 1. For Cog15, questions 27_Cog15 and 29_Cog15 appeared to be more challenging ($M < 0.50$) than the remaining items. The histogram displaying the score distribution for Cog15 in Figure B1 showed a mean of M = 10.19 and SD = 2.63. Additionally, reliability analyses (see Table 2) indicated that the measures were subthreshold for the items of the Cog15, with Cronbach's $\alpha = .67$, which would be considered questionable (Cronbach, 1951), and McDonald's $\omega = .70$, which also classifies as subpar (McDonald, 1999).

For ICAR16, the last four items of the test appeared more difficult to answer ($M < 0.50$). The histogram displaying the score distribution for ICAR16 in Figure B2 showed a mean of M = 9.22 and SD = 3.36. Furthermore, following reliability analyses (see Table 3), the remaining items of ICAR16 returned Cronbach's $\alpha = .75$ and McDonald's $\omega = .75$, both deemed acceptable values for the scale (Cronbach, 1951; McDonald, 1999).

Table 1

Descriptive statistics for Cog15 and ICAR16 scales

Cog15	
Item	Mean
2_Cog15	0.73
4_Cog15	0.89
7_Cog15	0.89
9_Cog15	0.50
11_Cog15	0.83
13_Cog15	0.94
14_Cog15	0.70
19_Cog15	0.86
21_Cog15	0.74
22_Cog15	0.62
23_Cog15	0.66
25_Cog15	0.51
26_Cog15	0.71
27_Cog15	0.30
29_Cog15	0.27
ICAR16	
Item	Mean
1_ICAR16	0.78
2_ICAR16	0.66
3_ICAR16	0.64
4_ICAR16	0.88
5_ICAR16	0.75
6_ICAR16	0.72
7_ICAR16	0.64
8_ICAR16	0.45

9_ICAR16	0.61
10_ICAR16	0.67
11_ICAR16	0.70
12_ICAR16	0.53
13_ICAR16	0.21
14_ICAR16	0.29
15_ICAR16	0.40
16_ICAR16	0.27

N = 428

Table 2

Alpha and omega values for Cog15

	Cronbach's α	McDonald's ω	Items
Cog15[1]	.71	—	15
Cog15[2]	.67	.70	15
Cog15[3]	.67	.69	10

[1]Original Cog15 from Kajonius (2014).

[2]Original Cog15 with sample (N = 428) from current study.

[3]Final version of Cog15 obtained in CFA from current study.

Table 3

Alpha and omega values for ICAR16

	Cronbach's α	McDonald's ω	Items
ICAR16[1]	.81	.83	16
ICAR16[2]	.75	.75	16
ICAR16[3]	.78	.78	12

[1]Original ICAR16 from Condon & Revelle (2014).

[2]Original ICAR16 with sample (N = 428) from current study.

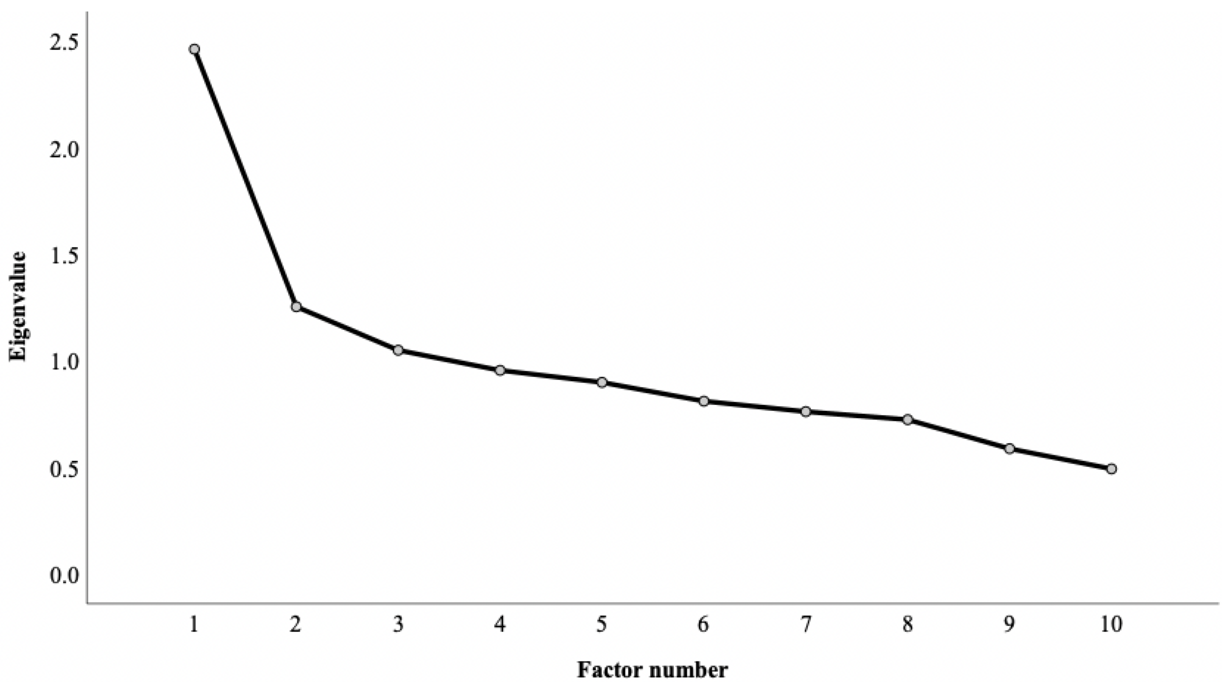
[3]Final version of ICAR16 obtained in CFA from current study.

Cog15: PCA and CFA

To see how well the Cog15 scale captures the g-factor, a series of PCAs were run on the first half of the dataset. The initial PCA showed that Bartlett's test of sphericity was significant (χ^2 (45, 213) = 199.98, $p < 0.001$), confirming that the model was a good fit for a factor analysis (Field, 2000). Additionally, based on Kaiser-Meyer-Olkin's measure of sampling adequacy (KMO = .706), it was deemed acceptable to proceed with the analysis (Field, 2000). Following the PCAs performed, items 2, 11, 19, 22, and 29 were removed from the Cog15, due to having loadings lower than .32 (Tabachnick & Fidell, 2007). According to the criterion of eigenvalue > 1, and the scree plot obtained (see Figure 1), the remaining Cog15 items were grouped into one factor explaining 24.62% of the total variance.

Figure 1

Scree plot from the PCA on Cog15



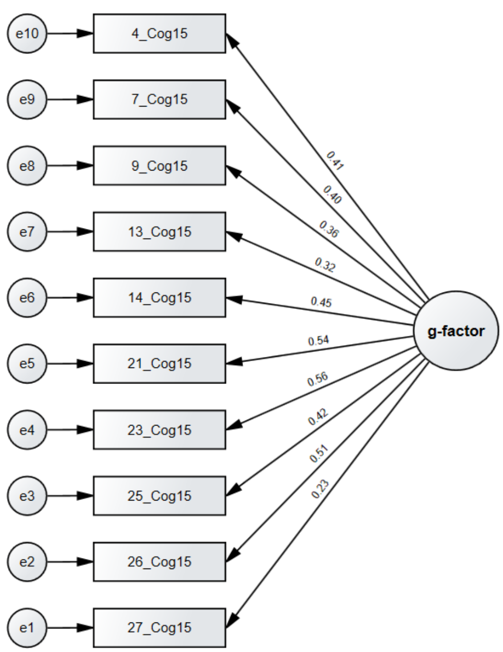
Note. Small circles = factors. We are only looking at the first factor with an eigenvalue of 2.5. We can see a drastic drop in the curve from the first factor to the second, which indicates that a single factor solution is an appropriate way to go. The single factor represents the g-factor.

In the final model, all remaining items had significant loadings above .32 (see Table C1), ranging from .35 to .71 (*M* = 0.48, *SD* = 0.12). Item 23 had the highest factor loading (.71) and item 7 had the lowest factor loading (.35).

Second, a CFA was conducted on the other half of the dataset for Cog15 to confirm the results returned from the previous PCA. All remaining items from the final PCA were grouped into one factor. Overall, based on the different model indices, the CFA indicated that the model showed a good fit to the data; RMSEA = .03, 90% CI [.00, .06]; SRMR = .05, CFI = .96. In contrast, the model showed poor fit on the chi-square test; $\chi 2$ (35,215) = 43.20, $p > 0.01$. All factor loadings were above .20 (see Table C1), ranging between .23 and .56 (*M* = 0.42, *SD* = 0.10), and were significant with $p < .001$. Item 23 had the highest factor loading (.56) and item 27 had the lowest factor loading (.23). Table 2 displays the new alpha and omega values obtained from the remaining items of the Cog15 scale. The path diagram produced from the CFA further details the g-factor, Cog15 items, and their standardized loadings (see Figure 2).

Figure 2

Path diagram for the CFA conducted on Cog15



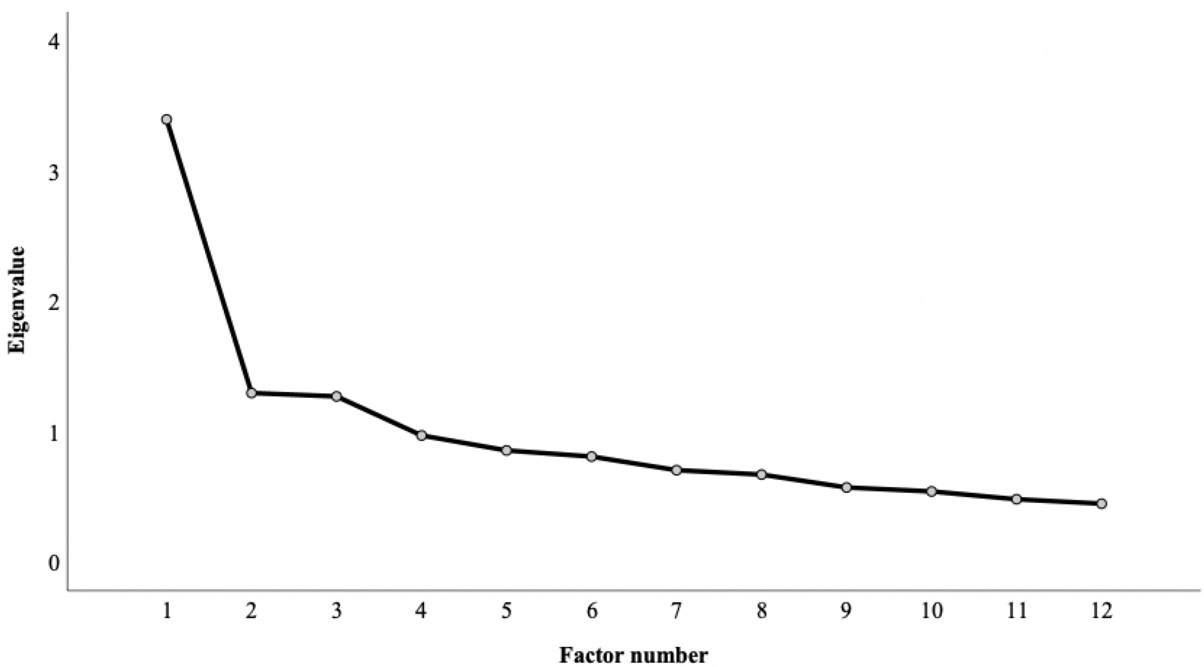
Note. Small circles = unobserved variables, rectangles = survey items (observed variables directly contributing to the factor), big circle = the single factor (g-factor), arrows = correlations between survey items and the single factor (g-factor), number on arrow = factor loading

ICAR16: PCA and CFA

First, a series of PCAs were run on the first half of the dataset for the ICAR16 to see how well the scale captures the g-factor. Bartlett's test of sphericity being significant (χ^2 (66, 213) = 408.11, $p < 0.001$), in addition to the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO = .788), both confirmed that the model was a good fit for factor analysis (Field, 2000). Following the PCAs performed, items 3, 4, 5, and 11 were removed from the ICAR16, due to having loadings lower than .32 (Tabachnick & Fidell, 2007). According to the criterion of eigenvalue > 1, and the scree plot obtained (see Figure 3), the remaining ICAR16 items were also grouped into one factor explaining 28.30% of the total variance.

Figure 3

Scree plot from the PCA on ICAR16



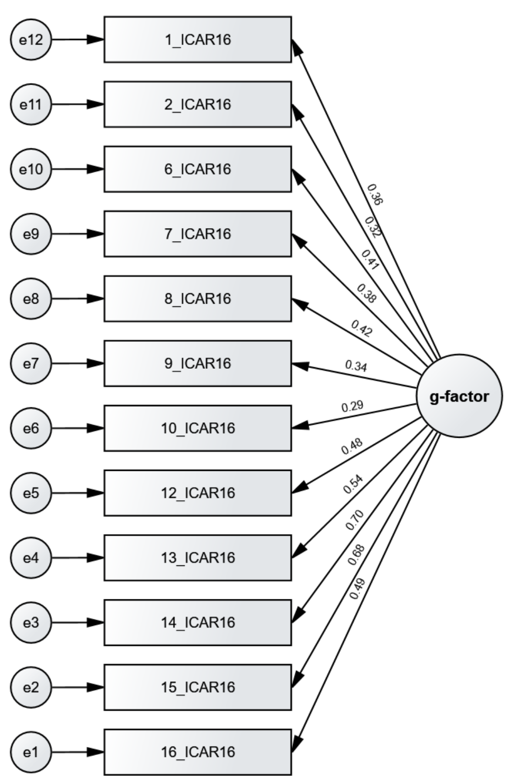
Note. Small circles = factors. We are only looking at the first factor with an eigenvalue of 3.4. We can see a drastic drop in the curve from the first factor to the second, which indicates that a single factor solution is an appropriate way to go. The single factor represents the g-factor.

In the final model, all remaining items had significant loadings above .32 (see Table C2), ranging from .41 to .59 (M = 0.53, SD = 0.06). Item 14 had the highest factor loading (.59) and item 12 had the lowest factor loading (.41).

Second, a CFA was conducted on the other half of the dataset to confirm the results obtained by the PCA. All items that were conserved in the final PCA were added into one factor. Overall, based on the different model indices, the model showed a poor fit to the data; RMSEA = .11, 90% CI [.10, .13]; SRMR = .09, CFI = .68. In contrast, the model showed a good fit on the chi-square test; $\chi 2$ (54,215) = 204, $p < 0.01$. All factor loadings were above .20 (see Table C2), ranging between .29 and .70 (M = 0.45, SD = 0.13). Item 14 had the highest factor loading (.70) and item 10 had the lowest factor loading (.29). Table 3 displays the new alpha and omega values obtained from the remaining items of the Cog15 scale. The path diagram produced from the CFA further details the g-factor, ICAR16 items, and their standardized loadings (see Figure 4).

Figure 4

Path diagram for the CFA conducted on ICAR16



Note. Small circles = unobserved variables, rectangles = survey items (observed variables directly contributing to the factor), big circle = the single factor (g-factor), arrows = correlations between survey items and the single factor (g-factor), number on arrow = factor loading.

Discussion

The aims of the present study were to examine the psychometric properties of two public domain intelligence tests, the previously untested Cog15, and the ICAR16, and to compare which of the two best captures the g-factor. The motivation behind these research aims was the need for more free resources to test intelligence, as the currently available options can be costly (Bainbridge, 2020). The analysis showed that ICAR16 was a superior tool for capturing the g-factor since it explained more of the variance in the data and had higher reliability scores. In the following, the reported results of the PCA, CFA, and reliability analyses performed on the two scales mentioned will be discussed.

How did Cog15 perform?

As indicated by the PCA, some items had to be dropped, more specifically 2, 11, 19, 22, and 29. A CFA was conducted on the remaining items, and the model fit was satisfactory. Question 27 had a factor loading lower than .32 and should be dropped. The scale's psychometric properties, Cronbach's alpha, and McDonald's omega were questionable, which can reflect poor quality in the dataset. Since the Cog15 scale is new and previously untested, further research on a larger sample is needed to see if these items should be dropped from the scale.

How did ICAR16 perform?

Following the PCAs conducted on ICAR16, four items from the original 16 were omitted (see Appendix A): 3, 4, 5 (Verbal Reasoning) and 11 (Matrix Reasoning). A point to consider is that all but one item from the original Verbal Reasoning factor in ICAR16 were dropped following the PCA conducted (Condon & Revelle, 2014), which begs the question of why the one remaining item loaded significantly onto the general factor. This could be due to the given sample at hand, however it is unclear why these items did not perform and further research is needed. We performed a CFA on the remaining items and the model fit indices (RMSEA, SRMR, CFI) indicated a poor fit for the data, yet the chi-square test indicated a good fit. One reason for the poor model fit indices could have been due to some of the indices being sensitive to categorical variables (Montoya & Edwards, 2021). Additionally, item 10 returned a factor loading below .32, suggesting it should also be dropped from ICAR16 (Tabachnick & Fidell, 2007).

Cog15 or ICAR16: Which scale captures the g-factor better?

Our results showed a single factor model for the Cog15 scale, which is not in line with Kajonius' (2014) original paper, which stated Cog15 aims to measure four different factors; numerical series, pattern reasoning, spatial ability, and verbal skills. Again, since Cog15 is a new and previously untested scale, we have limited prior research to compare our results with, and further research is needed. Our ICAR16 analysis identified one general factor, which is in line with previous research conducted by Kirkegaard & Nordbjerg (2015) on a Danish sample.

Despite these mixed results, it appears that, in this research paper, ICAR16 outperforms Cog15 in most aspects. ICAR16 explained more of the variance in the dataset than Cog15, which suggests that it would be better for measuring the g-factor. Moreover, ICAR16 had larger factor loadings on average (M = 0.45) in comparison to Cog15 (M = 0.42), which would suggest that the items included in ICAR16 contributed more to the g-factor than the items included in Cog15 (Yong & Pearce, 2013). ICAR16 also outperformed Cog15 in terms of reliability analyses, given that it returned higher alpha and omega values compared to Cog15. The values returned for ICAR16 are both deemed acceptable, while those obtained for Cog15 are considered questionable (Cronbach, 1951; McDonald, 1999). Overall, given these results, it seems that ICAR16 is a better measure of the g-factor than Cog15.

The g-factor is important to research and develop because it is a stable structure which changes little over time (Bock et al., 2000). Additionally, it is also a hereditary construct, it is genetically explained and contributes to the variance in intelligence test scores (De La Fuente et al., 2021). This would explain the need to understand whether the Cog15 and ICAR16 capture the g-factor or not, given that the two scales are open-access and available to measure intelligence. As previously mentioned, intelligence research can be costly due to how expensive it can be to administer intelligence tests (Bainbridge, 2020), and that can hinder research progress. It would benefit future researchers to have free resources that can measure a structure as important as the g-factor. Expanding our knowledge on publicly accessible intelligence tests could help future research in producing studies at a quicker and more efficient pace.

Limitations and Further Research

Despite ICAR16 appearing to be the better measure of the g-factor, some discrepancies present themselves in this paper's findings. One main limitation is that there is a gender imbalance within the sample, given that it is predominantly female. This limits the sample's diversity, and in turn, the generalizability of the results obtained. The lack of diversity in the sample could have affected the way some items performed, which could have led to them being dropped, in addition to potentially affecting the reliability of the scales. However, it is worth noting that the sample includes participants from different educational and occupational background, which diversifies the sample. Another limitation is that the items used in the questionnaire from the Cog15 scale, were all measuring pattern reasoning. ICAR16 had more diverse items which targeted different skills, while the questions in Cog15 appeared to measure the same concept. The questionnaire contained 50 items in total, it is possible that survey fatigue may occur due to the length of the survey (Porter, 2004). The questionnaire also did not contain any control items, to ensure participants remained engaged. This can further affect the results, especially if participants were experiencing survey fatigue (Porter, 2004). Another limitation could be due to external environmental factors since the test was not administered in a controlled setting. This could have had an impact on participants' attention and motivation while answering the survey.

Finally, one of the major limitations of the study was the binary dataset and the way that could have impacted the results of the factor analyses performed, both for the PCA and CFA. While the Bartlett test and KMO measures indicated that the data was fit to undergo factor analysis, the binary nature of the data (correct/incorrect answers) could have impacted the results of the PCA and CFA, in the sense of returning some inaccurate parameters measures (Silvia et al., 2012). However, given that the data was secondary, and previously collected over a timeline of several years, the only way to proceed was to recode the values into numerical data, which was done, as recommended by Silvia et al. (2012). Additionally, the maxiumum likelihood estimation methods and oblique rotations were also chosen to further ensure that errors could be avoided (Silvia et al., 2012). Additionally, previous research done on ICAR16 has also conducted factor analysis on the data despite having binary data in the form of correct or incorrect answers, and obtained significant and useful results in the process (Condon & Revelle, 2014; Keith & Young, 2020).

Future research should address these limitations mentioned above. Furthermore, researchers should replicate the results of this study on a larger, more inclusive, and diverse sample. This could aid future researchers in understanding Cog15 and ICAR16 better, as well as consolidating how well the two scales capture the g-factor, since previous samples on which ICAR16 was tested were smaller, and less diverse (Kirkegaard & Nordbjerg, 2015; Young & Keith, 2020). Adding control items to the scales and giving the participants a specific time frame to complete the test could also help future researchers in having more consistent and accurate results in their samples. Future researchers might benefit from administering the survey in a more controlled setting in comparison. Finally, proceeding with fewer items in Cog15 and ICAR16 as the results of the current paper indicated would tackle the possible problem of questionnaire fatigue (Porter, 2004).

Conclusion

Even though the present study has its limitations, it still contributes knowledge on the previously untested Cog15 scale and the ICAR16 scale. We established that ICAR16 is a better measure of the g-factor than the Cog15, based on the variance explained and reliability measurements. Since current intelligence tests can be expensive (Bainbridge, 2020), we believe it is important to research and validate these two free public domain intelligence tests, Cog15 and ICAR16, for future use.

Declaration of interest statement

The authors declare no potential competing interest.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101

Arbuckle, J. L. (2022). Amos (Version 29.0) [Computer Program]. Chicago: IBM SPSS.

Bain, S. K., & Allin, J. D. (2005). Book review: Stanford-Binet intelligence scales, fifth edition. *Journal of Psychoeducational Assessment*, *23*, 87–95.

Bainbridge, C. (2020, June 6). *Things you should consider before testing your child for giftedness*. Verywell Family. Retrieved February 9, 2023, from https://www.verywellfamily.com/testing-your-gifted-child-the-least-you-should-know-1449359

Bartlett, M. (1954). A note on the multiplying factors for various chi square approximations. *Journal of Royal Statistical Society*, *16*, 296–298.

Bock, G., Goode, J. A., & Webb, K. (2000). The nature of intelligence. In *Novartis Foundation Symposium*. Wiley. https://doi.org/10.1002/0470870850

Brody, N. (1999). What is intelligence? *International Review of Psychiatry*, *11*(1), 19–25. https://doi.org/10.1080/09540269974483

Brown, T. A. (2015). *Confirmatory factor analysis for applied research, Second Edition*. Guilford Publications.

Canivez, G. L. (2008). Orthogonal higher order factor structure of the Stanford-Binet Intelligence Scales--fifth edition for children and adolescents. *School Psychology Quarterly*, *23*(4), 533–541. https://doi.org/10.1037/a0012884

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge University Press.

Colom, R., Karama, S., Jung, R. E., & Haier, R. J. (2010). Human intelligence and brain networks. *Dialogues in Clinical Neuroscience*, *12*(4), 489–501. https://doi.org/10.31887/dcns.2010.12.4/rcolom

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

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/bf02310555

Deary, I. J. (2001). Human intelligence differences: a recent history. *Trends in Cognitive Sciences*, *5*(3), 127–130. https://doi.org/10.1016/s1364-6613(00)01621-1

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews Neuroscience*, *11*(3), 201–211. https://doi.org/10.1038/nrn2793

De La Fuente, J., Davies, G., Grotzinger, A. D., Tucker-Drob, E. M., & Deary, I. J. (2021). A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nature Human Behaviour*, *5*(1), 49–58. https://doi.org/10.1038/s41562-020-00936-2

DiStefano, C., & Hess, B. T. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment*, *23*(3), 225–241. https://doi.org/10.1177/073428290502300303

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989x.4.3.272

Field, A. P. (2000). *Discovering Statistics Using SPSS for Windows: Advanced Techniques for the Beginner*. SAGE.

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research*, *48*(5), 639–662. https://doi.org/10.1080/00273171.2013.804398

IBM Corp. (2022). IBM SPSS Statistics for Windows (Version 29.0). Armonk, NY: IBM Corp

Jensen, A. (1992). Spearman's hypothesis: Methodology and evidence. *Multivariate Behavioral Research*, *27*(2), 225–233. https://doi.org/10.1207/s15327906mbr2702_5

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/bf02291575

Kajonius, P. J. (2014). Honesty–Humility in contemporary students: Manipulations of self-image by inflated IQ estimations. *Psychological Reports*, *115*(1), 311–325. https://doi.org/10.2466/17.04.pr0.115c13z8

Kaufman, A. S., Flanagan, D. P., Alfonso, V. C., & Mascolo, J. T. (2006). Test review: Wechsler intelligence scale for children, fourth edition (WISC-IV). *Journal of Psychoeducational Assessment*, *24*(3), 278–295. https://doi.org/10.1177/0734282906288389

Kirkegaard, E. O., & Nordbjerg, O. (2015). Validating a Danish translation of the international cognitive ability resource sample test and cognitive reflection test in a student sample. *Open Differential Psychology*, 1-8.

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, *27*(3), 151–177. https://doi.org/10.1080/1047840x.2016.1153946

Lassiter, K. S., Bell, N., Hutchinson, M. B., & Matthews, T. D. (2001). College student performance on the general ability measure for adults and the Wechsler intelligence scale for adults-third edition. *Psychology in the Schools*. https://doi.org/10.1002/1520-6807(200101)38:1

McDonald, R. P. (1999). *Test theory: A unified treatment*. Psychology Press.

Montoya, A. K., & Edwards, M. J. (2021). The Poor Fit of Model Fit for Selecting Number of Factors in Exploratory Factor Analysis for Scale Evaluation. *Educational and Psychological Measurement*, *81*(3), 413–440. https://doi.org/10.1177/0013164420942899

Osborne, J. A. (2015). What is rotating in exploratory factor analysis? *Practical Assessment, Research and Evaluation*, *20*(2), 1–7. https://doi.org/10.7275/hb2g-m060

Porter, S. R. (2004). *Overcoming Survey Research Problems: New Directions for Institutional Research*. Jossey-Bass.

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5–26. https://doi.org/10.1177/0013164412449831

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. *Journal of Personality Assessment*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Seashore, H., Wesman, A., & Doppelt, J. (1949). The standardization of the Wechsler intelligence scale for children. *The Psychological Corporation*.

Silverman, W., Miezejeski, C., Ryan, R., Zigman, W., Krinsky-McHale, S., & Urv, T. (2010). Stanford-Binet and WAIS IQ differences and their implications for adults with intellectual disability (aka mental retardation). *Intelligence*, *38*(2), 242–248. https://doi.org/10.1016/j.intell.2009.12.005

Silvia, P. J., Wigert, B., Reiter-Palmon, R., & Kaufman, J. C. (2012). Assessing creativity with self-report scales: A review and empirical evaluation. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(1), 19–34. https://doi.org/10.1037/a0024071

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

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn and Bacon.

Taris, T. W., Kessler, S. R., & Kelloway, E. K. (2021). Strategies addressing the limitations of cross-sectional designs in occupational health psychology: What they are good for (and what not). *Work & Stress*, *35*(1), 1–5. https://doi.org/10.1080/02678373.2021.1888561

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach’s alpha. *International Journal of Medical Education*, *2*, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

The jamovi project (2022). *jamovi* (Version 2.3) [Computer Software]. Retrieved from https://www.jamovi.org

Thurstone, L. L. (1938). *Primary mental abilities.* University of Chicago Press.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. https://doi.org/10.1007/bf02291170

Yong, A., & Pearce, S. (2013). A beginner’s guide to factor analysis: Focusing on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94. https://doi.org/10.20982/tqmp.09.2.p079

Young, S. R., & Keith, T. Z. (2020). An examination of the convergent validity of the ICAR16 and WAIS-IV. *Journal of Psychoeducational Assessment, 38*(8), 1052–1059. https://doi.org/10.1177/0734282920943455

Appendix A

Sex/gender

- Male
- Female

Age

____

Part 2

A bat (slagträ) and a ball together cost 1.10 dollar. The bat costs 1 dollar more than the ball. How much does the ball cost?

____

If it takes 5 minutes for 5 machines to make 5 products, how many minutes will it take for 100 machines to make 100 products?

____

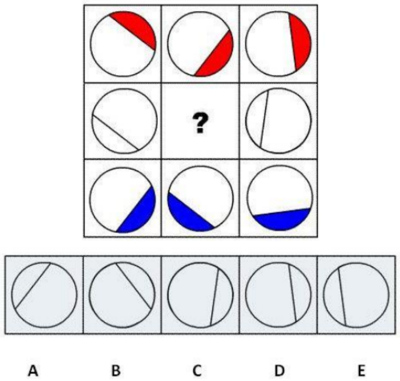
In a certain lake, water lilies (näckrosor) grow. Each day the area covered in the lake is doubled. It takes 48 days for the entire lake to be covered, so how many days does it take for half the lake to be covered?

____

Part 4 (Cog15)

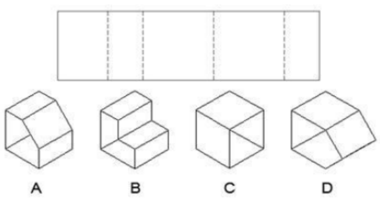
2.Which is the correct alternative for the question mark below (?)

1. A
2. B
3. C
4. D
5. E



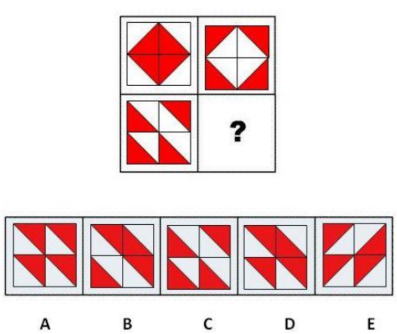
4.Which is the correct alternative for the box below?

1. A
2. B
3. C
4. D



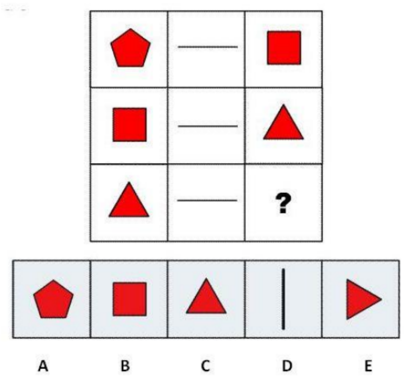
7.Which is the correct alternative for the question mark below (?)

1. A
2. B
3. C
4. D
5. E



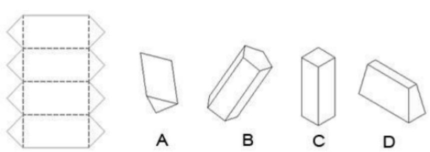
9.Which is the correct alternative for the question mark below (?)

1. A
2. B
3. C
4. D
5. E



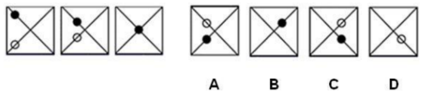
11.Which is the correct alternative for the folding below?

1. A
2. B
3. C
4. D



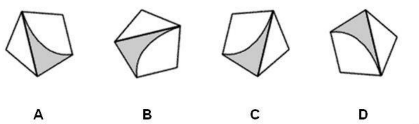
13.Which is the correct alternative for the movement below?

1. A
2. B
3. C
4. D



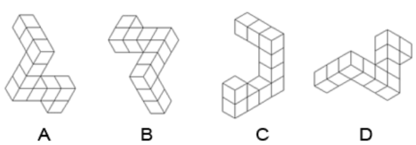
14.Which is the deviant alternative for the rotation below?

1. A
2. B
3. C
4. D



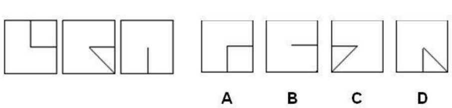
19.Which does not belong?

1. A
2. B
3. C
4. D



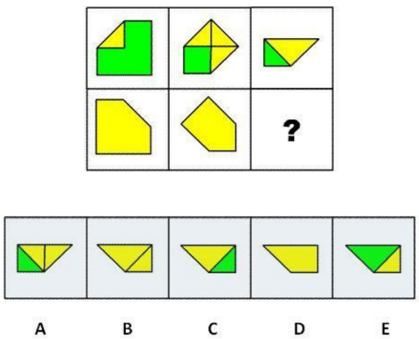
21.Which is the correct continuation below?

1. A
2. B
3. C
4. D



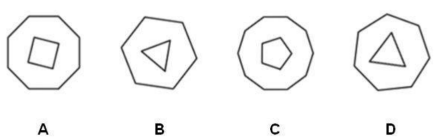
22.Which is the correct alternative for the question mark below (?)

1. A
2. B
3. C
4. D
5. E



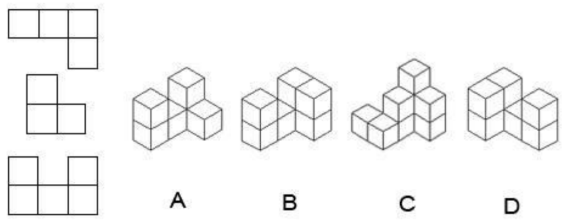
23.Which does not belong below?

1. A
2. B
3. C
4. D



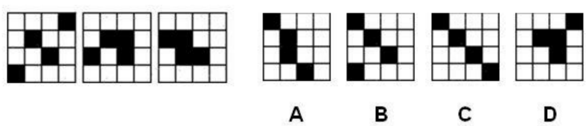
25.Which figure below matches the perspectives?

1. A
2. B
3. C
4. D



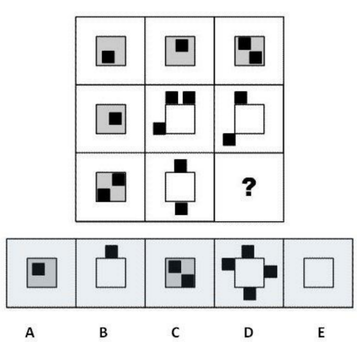
26.Which is the correct continuation below?

1. A
2. B
3. C
4. D



29.Which is the correct alternative for the question mark below (?)

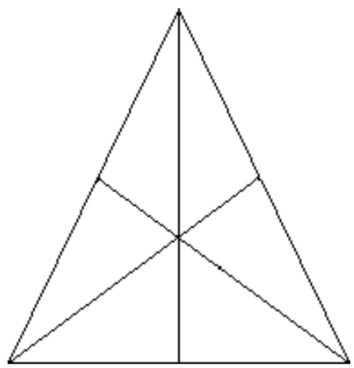
1. A
2. B
3. C
4. D
5. E



27. How many triangles in total can be extracted from the figure below?

1. 5-7
2. 8-9

3. 10-12
4. 13-15
5. >15



Part 5 (ICAR16)

1. What is one fifth of one fourth of one ninth of 900?

1. 2
2. 3
3. 4
4. 5
5. 6
6. 7
7. None of the above
8. I don't know

2. Which letter follows naturally? K, N, P, S, U, ...

1. S
2. T
3. U
4. V
5. W
6. X
7. None of the above
8. I don't know

3. Sakarias is taller than Matte, och Rikard is shorter than Sakarias. Which of the following is correct?

1. Rikard is taller than Matte
2. Rikard is shorter than Matte
3. Rikard is as tall as Matte
4. It is impossible to know
5. Rikard is taller than Sakarias
6. Sakarias is shorter than Matte
7. None of the above
8. I don't know

4. Joshua is 12 years old and his sister is 3 times as old as him. When Joshua is 23, how old will his sister be?

1. 35
2. 39
3. 44
4. 47
5. 53
6. 57
7. None of the above
8. I don't know

5. If the day after tomorrow is two days before Thursday, then what day is it today?

1. Friday
2. Monday
3. Wednesday
4. Saturday
5. Tuesday
6. Sunday
7. None of the above

8. I don't know

6. What letter comes next? V, Q, M, J, H, ...

1. E
2. F
3. G
4. H
5. I
6. J
7. None of the above
8. I don't know

7. What letter comes next? I, J, L, O, S, ...

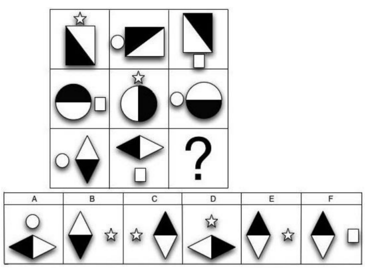
1. T
2. U
3. V
4. X
5. Y
6. Z
7. None of the above
8. I don't know

8. What letter comes next? Q, S, N, P, L, …

1. J
2. H
3. I
4. N
5. M
6. L
7. None of the above
8. I don't know

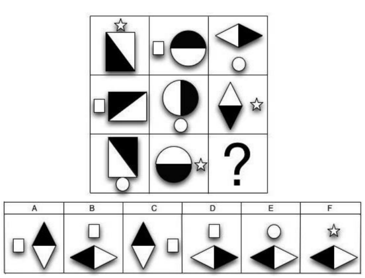
9. Which alternative fits best in the question mark (?)

1. A
2. B
3. C
4. D
5. E
6. F
7. None of the above
8. I don't know

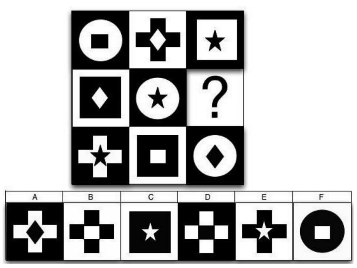


10. Which alternative fits best in the question mark below (?)

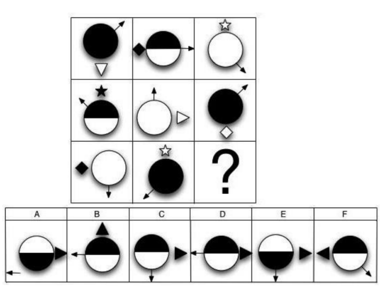
1. A
2. B
3. C
4. D
5. E
6. F
7. None of the above
8. I don't know



11. Which alternative fits best in the question mark below (?)
    1. A
    2. B
    3. C
    4. D
    5. E
    6. F
    7. None of the above
    8. I don't know

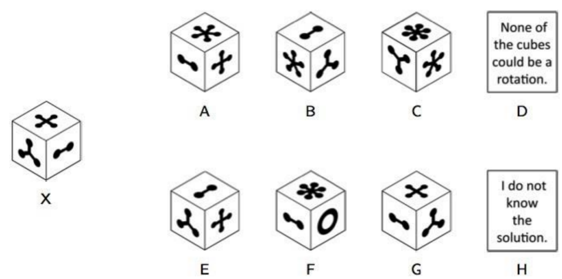


12. Which alternative fits best in the question mark below (?)
    1. A
    2. B
    3. C
    4. D
    5. E
    6. F
    7. None of the above
    8. I don't know



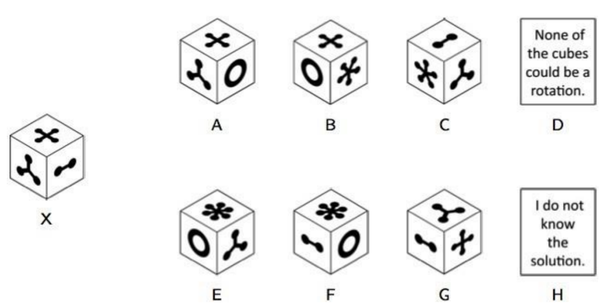
13. Which of these cubes below is a rotation of X?

1. A
2. B
3. C
4. D
5. E
6. F
7. G
8. H



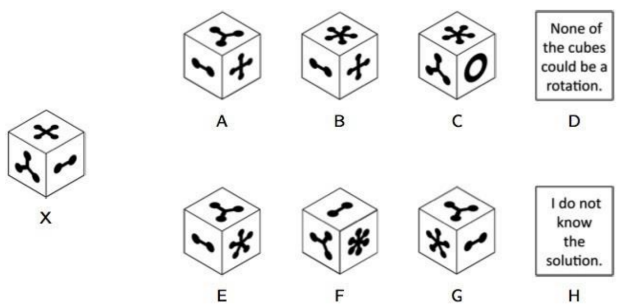
14. Which of these cubes below is a rotation of X?

    1. A
    2. B
    3. C
    4. D
    5. E
    6. F
    7. G
    8. H



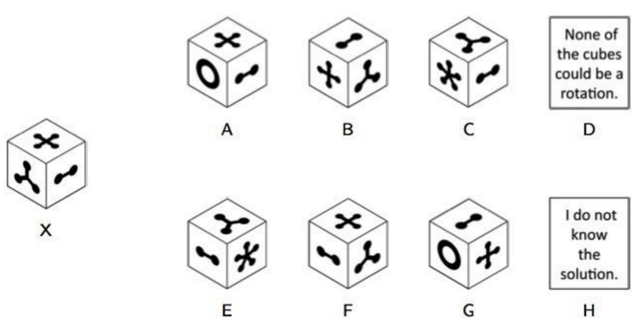
15. Which of these cubes below is a rotation of X?

1. A
2. B
3. C
4. D
5. E
6. F
7. G
8. H



16. Which of these cubes below is a rotation of X?

1. A
2. B
3. C
4. D
5. E
6. F
7. G
8. H



Correlations between gender, age, and the Cog15 items

Variable	Gender	Age	2_Cog15	4_Cog15	7_Cog15	9_Cog15	11_Cog15	13_Cog15	14_Cog15	19_Cog15	21_Cog15	22_Cog15
Gender	—											
Age	-.11*	—										
2_Cog15	.09	.07	—									
4_Cog15	.02	.07	.22**	—								
7_Cog15	.04	.00	.08	.20**	—							
9_Cog15	-.15**	-.01	.13**	.15**	.09	—						
11_Cog15	.11*	.02	.10*	.17**	.09	-.01	—					
13_Cog15	.00	-.04	.15**	.17**	.20**	.12*	.22**	—				
14_Cog15	-.09	-.01	.09	.12*	.07	.13**	.13*	.07	—			
19_Cog15	-.08	-.06	.08	.16**	.06	.05	.07	.15**	.02	—		
21_Cog15	-.11*	.05	.07	.20**	.12*	.23**	.14**	.14**	.24**	.02	—	
22_Cog15	.07	-.11*	.06	.05	.02	.06	.17**	.10*	.10	.14**	.11*	—
23_Cog15	-.06	.04	.16**	.18**	.23**	.30**	.16**	.23**	.21**	.11*	.28**	.05
25_Cog15	-.07	-.05	.05	.06	.09	.12*	.12*	.17**	.23**	.14**	.19**	.19**
26_Cog15	-.05	.01	.23**	.22**	.22**	.20**	.06	.29**	.20**	.15**	.18**	-.04
29_Cog15	.09	.03	-.05	.04	-.04	-.08	.07	.00	.01	-.13**	-.01	-.05
27_Cog15	-.04	.05	.06	.07	.08	.13**	.03	.00	.12*	.10*	.08	.18**

*. Correlation is significant at the 0.05 level (2-tailed).

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

Variable	Gender	Age	1_ICAR16	2_ICAR16	3_ICAR16	4_ICAR16	5_ICAR16	6_ICAR16	7_ICAR16	8_ICAR16	9_ICAR16	10_ICAR16	11_ICAR16
Gender	—												
Age	-.11*	—											
1_ICAR16	-.18**	.05	—										
2_ICAR16	.09	.00	.16**	—									
3_ICAR16	.02	.00	.08	.01	—								
4_ICAR16	.00	.00	.17**	.14**	-.01	—							
5_ICAR16	-.05	.07	.27**	.03	.06	.07	—						
6_ICAR16	-.01	.08	.32**	.25**	.06	.14**	.11*	—					
7_ICAR16	.03	-.05	.23**	.39**	.05	.13**	.13**	.35**	—				
8_ICAR16	-.03	.00	.28**	.27**	-.01	.15**	.06	.33**	.27**	—			
9_ICAR16	-.11*	.09	.20**	.10*	.03	.04	.09	.10*	.04	.26**	—		
10_ICAR16	-.04	.01	.20**	.10	.03	.07	.21**	.14**	.12*	.09	.33**	—	
11_ICAR16	.05	.00	.16**	.14**	-.03	.05	-.01	.16**	.15**	.22**	.19**	.18**	—
12_ICAR16	-.10	.07	.18**	.17**	.02	.10*	.12*	.24**	.14**	.21**	.23**	.28**	.23**
13_ICAR16	-.08	.02	.20**	.14**	.03	.11*	.13**	.15**	.16**	.18**	.12*	.10*	.08
14_ICAR16	-.04	.02	.21**	.16**	-.01	.11*	.12*	.18**	.19**	.19**	.13**	.18**	.07
15_ICAR16	-.08	.11*	.19**	.16**	.00	.07	.10*	.18**	.22**	.19**	.21**	.17**	.15**
16_ICAR16	-.06	-.01	.14**	.12*	.08	.08	.00	.09	.19**	.15**	.17**	.13*	.16**

*. Correlation is significant at the 0.05 level (2-tailed).

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

Figure B1

Histogram of total test scores for Cog15

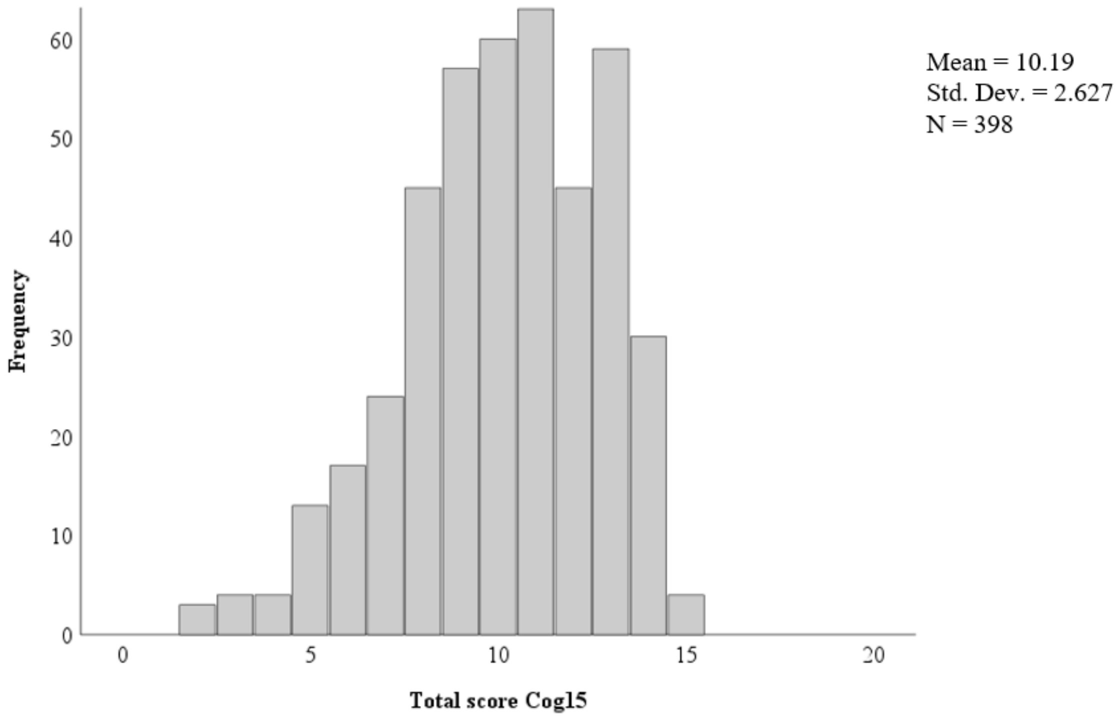
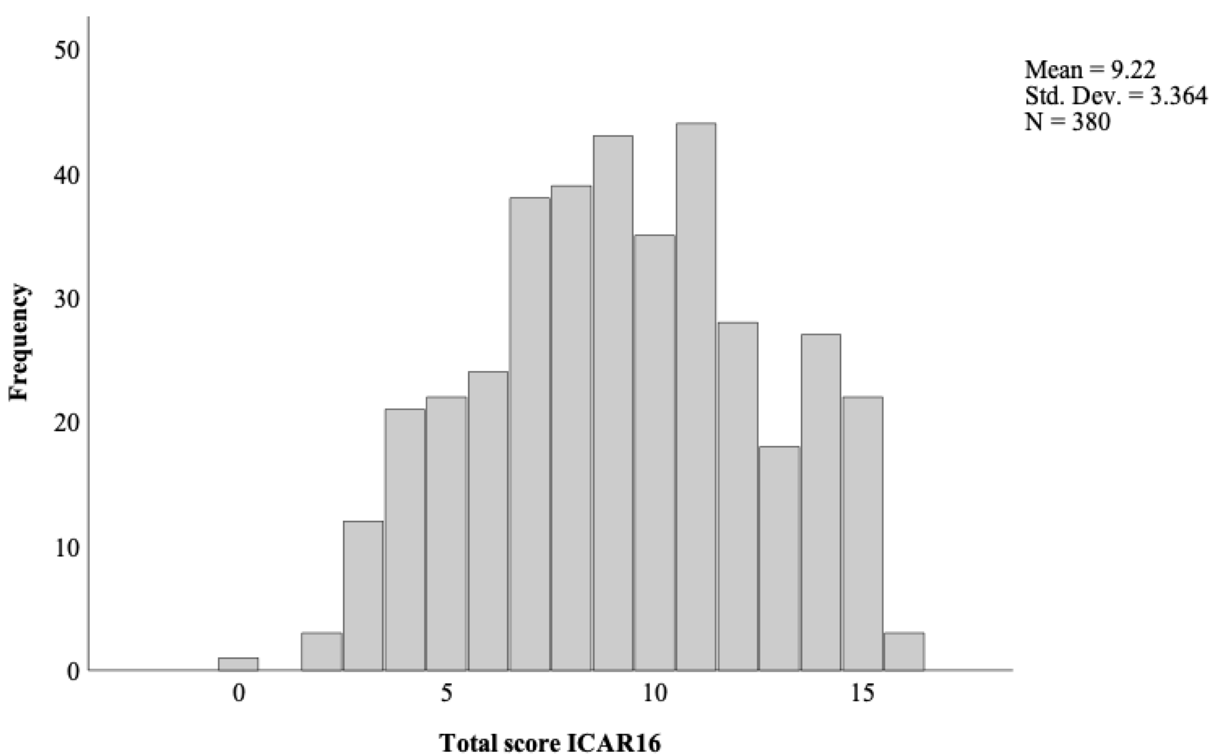


Figure B2

Histogram of total test scores for ICAR16



Appendix C

Table C1

Cog15 PCA and CFA loadings

Items	PCA (15 items)	PCA (10 items)	CFA (15 items)	CFA (10 items)
2_Cog15	*.25*		*0.43*	
4_Cog15	.48	.43	0.44	.41
7_Cog15	.32	.35	0.40	.40
9_Cog15	.50	.55	0.34	.36
11_Cog15	*.30*		*0.39*	
13_Cog15	.54	.50	0.35	.32
14_Cog15	.37	.41	0.46	.45
19_Cog15	*.31*		*0.28*	
21_Cog15	.42	.42	0.51	.54
22_Cog15	*.22*		*0.25*	
23_Cog15	.66	.71	0.58	.56
25_Cog15	.50	.50	0.41	.42
26_Cog15	.62	.62	0.47	.51
29_Cog15	*-.10*		*-0.08*	
27_Cog15	.35	.36	0.25	.23

Note. Direct oblimin rotation. Italicized items were removed due to PCA loadings below the threshold of .32 (Tabachnick & Fidell, 2007).

Table C2

ICAR16 PCA and CFA loadings

Items	PCA (16 items)	PCA (12 items)	CFA (16 items)	CFA (12 items)
1_ICAR16	.58	.56	.40	.36
2_ICAR16	.51	.53	.35	.32
3_ICAR16	*.08*		*.09*	
4_ICAR16	*.31*		*.17*	
5_ICAR16	*.29*		*.22*	

6_ICAR16	.52	.54	.46	.41
7_ICAR16	.57	.57	.41	.38
8_ICAR16	.54	.55	.46	.42
9_ICAR16	.41	.42	.37	.34
10_ICAR16	.45	.47	.33	.29
11_ICAR16	*.31*		*.31*	
12_ICAR16	.40	.41	.53	.48
13_ICAR16	.54	.55	.51	.54
14_ICAR16	.58	.59	.64	.70
15_ICAR16	.55	.56	.62	.68
16_ICAR16	.59	.59	.44	.49

Note. Direct oblimin rotation. Italicized items were removed due to PCA loadings below the threshold of .32 (Tabachnick & Fidell, 2007).