



RESEARCH ARTICLE

Psychometric properties of a brief non-verbal test of g factor intelligence

Luis Anunciacao¹, Louise Marques¹, Christopher Murray², Anna C. Portugal¹, Ivan Rabelo³, J. Landeira-Fernandez¹ and Roberto Cruz⁴

Published online: 30 May 2024.

Abstract

Intelligence is the most studied construct in psychology and cognitive neuroscience. In Brazil, the administration of intelligence tests is needed for a number of social rights, including driving privileges. Such requirements have led to a large testing industry but the vast majority of intelligence tests require extended administration times and language skills. In this study, we sought to investigate the psychometric properties and normative results of a new non-verbal intelligence test, the General Matrix of Intelligence (GMI). The GMI is comprised of 28 matrix-based items and can be administered in as little as six-minutes. In this initial pilot test, the GMI was administered to 1,326 participants, ages 15-64 years old ($M = 25.65$ years, $SD = 9.6$ years), from all regions in Brazil. These data were analyzed using a 2PL Item Response Theory model, regression analyses were conducted to determine the role of sociodemographic factors, and preliminary norms were computed. Results indicated a unidimensional solution that reproduced the g factor theory, invariance across genders, evidence that cognitively demanding items involving movement or three-dimensional shapes were more difficult than items with less cognitive load, a normal distribution for results, and an interaction between education level and age group in predicting performance. Implications of these findings for research and practice are discussed and all data and codes are provided at <https://osf.io/kvu42/>.

Keywords: g factor; general intelligence; invariance; intelligence assessment; psychometrics; Brazil.

INTRODUCTION

Intelligence is the most studied construct in psychology and cognitive neuroscience, predating the scientific methods that paved the origins of psychology in the mid-19th century (Gottfredson & Saklofske, 2009). Intelligence is often defined as the capacity to learn from experience, adapt to surrounding environments, and utilize resources to solve problems (Colom et al., 2010; Sternberg, 2018a). Due to its long history in psychological testing and its relationship with social debates (Hernstein & Murray, 2010; Weinberg, 1989), several other complementary theories of intelligence have been proposed including: (a) g factor and variations, such as Cattell-Horn-Carroll (CHC);

(b) multiple mental abilities (Gardner, 1983), and (c) emotional intelligence (Goleman, 1995). Prior findings, however, provide limited support for the emotional perspective, suggesting it is a successful attempt to rebrand personality (Waterhouse, 2006) and, although the results of factor analyses demonstrate that multiple intelligences do exist (Deary, 2012), they are better explained as nested functions in a hierarchical system in which g factor is the core.

Intelligence as a Construct

The seminal and first theory of intelligence was proposed in 1904 by Charles Spearman, labeled the g (general) factor theory (Spearman, 1904; Warne & Burningham, 2019). Some years later in the 1930s, Thurstone challenged this perspective and proposed the Primary Mental Abilities (PMAs) theory, claiming that intelligence was comprised of several different factors (Thurstone, 1938). From this perspective, seven primary skills including verbal comprehension, word fluency, number facility, spatial visualization, associative memory, perceptual speed, and reasoning were included. In the early 1960's, Cattell (1963) returned to the Spearman's general factor approach but argued that general

¹ Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro, Brazil

² Center on Human Development, University of Oregon, USA

³ Universidade Paulista (UNIP)

⁴ Federal University of Santa Catarina (UFSC)

*) *corresponding author*

Luis Anunciacao

Email: luisfca@puc-rio.br

intelligence should be divided into fluid (g_f) and crystallized intelligence (g_c). One of his doctoral students at the time, J. L. Horn, identified other broad intellectual abilities to supplement fluid and crystallized intelligence. In the following years, John Carroll conducted a factor analytic study of several cognitive and demographic variables from over 460 different datasets and suggested a hierarchical three-layered model comprised of a narrow, broad, and general cognitive ability layer. These last three theories are mainly based on the g factor perspective and due to their complementary aspects, tend to be named the CHC theory (Deary, 2012). To retain the multi-factor theory originally proposed by Thurstone, Gardner (1983) proposed a model that incorporated interpersonal and intrapersonal intelligence. These last two factors provided a preliminary foundation for moving from definitions of intelligence as purely cognitive, to definitions that incorporated a social- and emotional-basis for intelligence, a perspective further popularized by Goleman (1995).

Despite these multiple theories of intelligence, the g (general) remains the most widely accepted definition among specialists in the field (Rindermann et al., 2020) and continues to have the strongest basis in the empirical literature (Warne & Burningham, 2019). Initially proposed by Spearman when analyzing data from 22 boys in a preparatory school of the highest class (Horn & McArdle, 2007), Spearman's original "two-factor theory" led him to conclude the existence of an underlying g factor responsible for all intellectual activity, and a second factor (s) that is specific to the task. In identifying this model, Spearman essentially developed the first version of factor analysis methodology, which boosted links between psychology and statistics with intercorrelations between all ability tests referred to as "positive manifold" (Van Der Maas et al., 2006). A seminal study that relied on a data analysis of over 400 participants after they completed three cognitive batteries revealed that all results were strongly and positively correlated (i.e., a positive manifold effect), and each test battery had a strong g factor. These test batteries included 14 tests from the Hakstian and Cattell Comprehensive Ability Battery, 17 tests from the Hawaii Battery, and 11 tests from the Wechsler Adult Intelligence Scale (Deary, 2012). Since this study, the concept of "just one g " has been solidified (Johnson et al., 2004).

The g factor provided a mathematical basis for contemporary models of intelligence, in which g is a higher-order factor that accounts for observed covariance among broad ability factors (Conway et al., 2021). Modern procedures have not modified the rationale that underlies statistical analyses, but helped to increment, and generalize, their structure and mathematical accuracy (Carroll, 2003). Recent studies have demonstrated that Spearman's model also adequately fits data from non-Western countries (Warne & Burningham, 2019).

The development of psychological tests is the cornerstone of intelligence. When such tests have adequate psychometric properties, their results enable indirect descriptions of cognitive skills among test takers along with statistical summaries of different groups of people according to their demographic and personal characteristics (Koch et al., 2021). Despite controversies related to these tests within studies of human intelligence (Detterman, 2014; Hernstein & Murray, 2010), the results are nonetheless important. They are explicitly or implicitly considered predictors of various outcomes, including performance in professional fields, educational achievement, and other everyday settings (Bünger et al., 2021; Neisser et al., 1996).

Intelligence Testing in Brazil

Brazil has a long history of using psychological testing—including intelligence tests—for certain social rights including access to firearms and driving privileges. Even before the first regulations governing the professional practice of the nation's psychologists in 1962, legislators understood that mental health certificates were necessary for certain *prima facie* rights (Silva et al., 2009). For driving privileges, such procedures began in the 1940s. The following decades witnessed an expansionist economic policy and the automotive industry became one of the most important and strategic sectors in the Brazilian economy. Cars were consequently purchased at high rates, but increased automobile ownership was accompanied by increases in traffic accidents, thus prompting politicians and the public to initiate efforts to improve driver safety. Because automobile accidents were primarily attributed to human error, policy makers reasoned that having a better understanding of the psychological and health factors associated with these events could help to prevent them (Silva et al., 2009), a proposition that is now supported by recent evidence demonstrating positive associations between cognitive functioning and driving performance in complex situations (Anstey et al., 2009; Anstey et al., 2006).

As a result of policies requiring psychological testing for driving privileges in Brazil, a formal alliance between psychologists and policymakers was established providing the social and legal structure for what is now known as traffic psychology (Ciotta, Cruz, & Dagostin, 2021). Policies governing traffic psychology include a set of legal standards that define the necessary psychological attributes for driving privileges as well as required levels of performance for accessing driving privileges. Although standards were initially based on accident-proneness, over time these standards transitioned to a cognitive-based approach. Currently, every Brazilian citizen who wants to drive a motor vehicle is required to participate in a comprehensive evaluation that includes assessments of attention, memory, intelligence, and personality (Contran nº 927/2022). Thus, traffic psychologists have become an indispensable state workforce who provide summary reports of the cognitive abilities of individuals pursuing a driver's license and psychological testing is the primary approach for evaluating eligibility.

When conducting such testing, traffic psychologists are required to use psychology tests certified by the Brazilian Federal Council of Psychology. This public organization regulates the validity of psychological tests used in Brazil. However, as the system became more tightly controlled, several unintended consequences arose. First, the testing industry in Brazil tends to have a major focus on traffic, as it is closely linked with the process of obtaining a driver's license. This focus has led to the development and validation of numerous tests but has also contributed to a dearth of tests for other purposes, such as neuropsychological assessment causing many psychologists to rely on tests developed for driving privileges for other purposes (Anuniação et al., 2021). Second, due in part to the fact that intelligence testing for driving privileges is "high-stakes" (i.e., a requirement for driving) many of the existing tests used in the context of driving privileges have been leaked to the general public, raising concerns about the trustworthiness of test results. Third, the vast majority of the currently available tests are not designed to capture the speed of changes in the Brazilian traffic legislation, leading to problems with

normative reference for new demands that may arise due to modifications in legislation (Anuniação et al., 2021).

In addition to these challenges, additional challenges such as the length of time required for such testing and language requirements can also affect the efficiency and quality of testing procedures. For example, the majority of tests used to determine driving privileges require a large amount of time to administer. Because all citizens interested in pursuing driving privileges are required to participate in testing, implementing such testing is highly resource intensive. Moreover, and consistent with most intelligence testing, the majority of tests rely heavily on language skills. However, Brazil is a large diverse culture that includes over 100 languages, potentially disadvantaging individuals who do not speak Portuguese as a first language. Similarly, individuals with various learning needs such as those with language or learning disabilities may also experience challenges when completing tests that rely heavily on language skills.

Current Study

The purpose of the current project was to conduct an initial pilot test of a brief, non-verbal intelligence measure for use in Brazil. Because intelligence testing plays a crucial role in various aspects of individual and societal functioning within Brazil, it is important to continue to develop new measures that can be used to assess intelligence efficiently and effectively. Such measures should be relatively easy to implement, they should allow for testing in diverse populations, and they should possess strong psychometric properties including updated norms. In this study, we evaluated a measure that was designed in non-verbal format—the General Matrix of Intelligence (GMI)—to expand the repertoire of reliable and valid assessment tools available for addressing a diverse range of individuals, including those with language barriers or cultural differences (Anastasi & Urbina, 1997; Flanagan et al., 2012). In doing so, we sought to evaluate the structural properties of the GMI, we conduct invariance testing across gender, and we evaluate normative data based on means and IQ score conversions. We considered this research exploratory in the sense that this study represents an initial effort to develop and evaluate the psychometric properties of the GMI in Brazil.

METHODS

Participants

Prior to recruiting any participants, this study was approved by the Brazilian Institutional Review Board (public note *hidden for peer-review*). A total of 1,326 Brazilian participants comprised the study sample. The participants were located in all Brazilian regions but mostly in the southeast ($n = 1,106$ [83.7%]), northeast ($n = 133$ [10.1%]), and south ($n = 83$ [6.3%]). Their ages ranged from 15 to 64 years ($M = 25.65$ years, $SD = 9.6$ years), with 57.5% ($n = 763$) female. Highest level of education included elementary school ($n = 63$ [4.9%]), high school ($n = 427$ [33.3%]), and college undergraduate level and above ($n = 792$ [61.8%]).

Instrument

The GMI is a non-verbal intelligence test, developed to be matrix-based and respondents must think abstractly to

perceive and identify relationships in geometric figures. In Figure 1 we provide an example of how items were structured. The test is based on “noegenetic laws or principles” conceived by Spearman to assess the g factor. For Spearman, these principals refer to the capacity to create, understand, and build knowledge based on what is sensed, perceived, and comprehended. Three noegenetic laws account for this capacity: (a) the law of apprehension of experience (i.e., ability to perceive fundamental features of a given problem), (b) the law of eduction of relations (i.e., ability to discover links between two or more ideas or schemas), and (c) the law of eduction of correlates (i.e., ability to extrapolate, generalize, and build a new idea to infer a not-immediately-educed relation from extant relations, Horn & McArdle, 2007).

The GMI was developed using the following procedures. First, a set of 40 familiar items was initially developed, in which such elements as the format of the geometric figure (two- or three-dimensional), its color, its quantity, its movement, and its mathematical relationship were present. This initial version was administered to a group of 30 undergraduate students to check whether the items were extremely easy or difficult and to evaluate unintentional similarities among items, or procedures that could lead to bias. Results of the first administration allowed us to generate a second version that consisted of 30 items. The second version of the measure was administered to a new group of participants to determine how much time was necessary to implement the GMI and to check item quality. After completing this step, two items were removed on the basis of descriptive statistics, resulting in a new 28-item version, and six minutes was defined as the time limit for implementing the test. The set of 28 items was then examined by four independent experts in psychology and assessment, all of whom held PhD or PsyD degrees with a record of contributions to the field. These reviewers were asked to provide suggestions for changing items, instructions, or other features to improve overall test quality. Based on expert review, minor changes to item presentation order were made. The scoring system for the final GMI is dichotomous (i.e., correct/incorrect). Thus, the total score can range from 0 (i.e., the participant did not correctly answer any items) to 28 (i.e., correctly answered all items).

Procedures

Brazilian driver's licenses are conditional on medical and psychological assessments, resulting in traffic clinics throughout the country. In our study, we took advantage of this requirement for recruiting participants from clinics during their default assessment for obtaining a driver license. Test administration was conducted by test providers in small groups (≤ 10 participants per group) between 2013 and 2020. Because participants were in the process of obtaining their driver's license, the test was piloted as part of a set of approximately four other measures.

All tests were administered by trained tests administrators who received a free copy of the test, formal training by the authors before administration, and a small incentive for administering the test (i.e., R\$ 20). Trainings lasted about one hour, and all test providers could ask clarifying questions and request guidance pertaining to testing procedures. All registration numbers of the test providers were verified to ensure qualifications.

All study participants were informed of the goal of the assessment, they received a consent form, were informed that confidentiality would apply to their involvement, and

they were provided an opportunity to ask questions about the study prior to testing. After receiving participant assent, the testing procedure was administered.

Analytic Approach

All data were examined using tables and graphs to detect computational errors and incorrect coding. Outliers were not suppressed, and missing cases of the test data were not present. For data processing, multiple Excel spreadsheets were harmonized in R programming language. All analyses were performed in R 4.0, with support from the *mirt* (Chalmers, 2012) and *lavaan* (Rosseel, 2012) packages. Codes and R notebooks are freely available at *[hidden for peer-review]*. Analyses were performed using the 2-parameter logistic (2PL) Item Response Theory (IRT) and Categorical Confirmatory Factor Analysis (CFA).

IRT models comprise a set of statistical techniques that deal with the interaction between a participant's ability and item parameters. The 2PL model allows items to vary by both difficulty (*b* parameter) and discrimination (*a* parameter), and the CFA produces goodness-of-fit based on the weighted least square mean estimator. In IRT, the *b* parameter is a threshold that indicates the location where the probability of success is .5, and the *a* parameter is the slope of the item characteristic curve at the point $\theta = b$ to describe how well an item can differentiate between examinees having abilities above or below the item location. Steeper curves indicate better item discrimination (Baker & Kim, 2017). The first model was computed with the Ramsay algorithm for expectation-maximization (EM) acceleration and the Gaussian function of the latent density type. The fit of both models was evaluated via either M^2 or χ^2 statistics, along with the values of the comparative fit index (CFI), Tucker-Lewis index (TLI), and root mean square error of approximation (RMSEA).

Measurement invariance is a fundamental property of a test and ensures that individuals that have the same score on a latent scale and provide similar responses on observed indicators (i.e., test items). Indicators of change, such as ΔCFI , ΔTLI , and $\Delta RMSEA$, were used to determine whether the equivalence of pattern (configural or form), loadings (weak or metric), and intercepts (strong or scalar) should be retained. The cutoffs that were proposed by Chen (2007) were implemented. Reliability analyses were performed within IRT framework and using Classical Test Theory (CTT). In IRT, as reliability varies across the range of theta, we provide the test information curves. In CTT, we checked the internal consistency with Cronbach's alpha and Guttman's lambda.

We investigated outliers in test results by using the Grubbs' test. This method is useful for detecting outliers in a univariate normal distribution and its null hypothesis states that there are no outliers in the data. We then used linear methods to check the effects of demographic variables (e.g., sex and level of education) on test results. Post hoc analyses were performed with the Tukey method for adjusting p-values estimates. Partial eta squared was used to compute effect sizes. This effect size considers the variance explained by a given variable of the variance remaining after excluding variance explained by other predictors. Its interpretation should be .01 to indicate a small effect; .06 to indicate a medium effect; and .14 to indicate a large effect.

Preliminary normative references were computed based on statistical significance and effect size results. To perform these analyses, we relied on the final raw score of the test (i.e., the summative score of all 28 items). The cumulative frequency of these results was computed and

transformed into percentiles. The normal curve z-score equivalent for each percentile rank was then computed and fitted into a theoretical IQ distribution, with a mean of 100 and standard deviation of 15. Finally, the raw data results were linked to the theoretical normal distribution to scale each raw result into an IQ score. For example, based on the data presented here, a score of 9 points equated to a percentile rank 16%. In a standard normal curve (z-score distribution), the percentile of 16% represents 1 standard deviation below the mean. Thus, assuming a theoretical mean and standard deviation of 100 and 15 respectively, everyone who scored 9 points in the test would be assigned an IQ of 85. The comparison of the performance of this transformation was checked by computing the empirical z-scores and checking the difference between this empirical result and the theoretical standard normal z (Burns, 1988).

RESULTS

Descriptive statistics, CFA results, and IRT estimates for the 28 GMI items are presented in **Error! Reference source not found.** The average proportion of correct responses was 54%. The unidimensional solution achieved an adequate goodness-of-fit in the CFA and IRT analyses. The first was $\chi^2(350) = 5767.375$, $p < .001$, CFI = .935, TLI = .930, RMSEA = .091, and the latter was $M^2(350) = 2864$, $p < .001$, CFI = .915, TLI = .909, RMSEA = .074. Factor loadings are provided in Figure 2. The test information for θ (or ability) ranging within -1 +1 was 4.94 and 16.1. The peak of information was 21 when $\theta = 0$ (average). Classical test reliability analyses were adequate and included a coefficient alpha of .879 and Guttman's lambda of .9. These results make it possible to use the GMI's scores as for proxy of non-verbal general intelligence.

The IRT analyses allowed us to conclude that item 20 was the most discriminative as its curve was steeper than others (see Figure 3). Therefore, it had the greatest ability to detect subtle differences in the respondents' abilities. The probability of correctly replying to this item was approximately 20% for those with an average *a* trait level. However, almost all individuals who were slightly above average would correctly respond to this item, whereas the other individuals would fail this item. In turn, item 5 was the least discriminative (see Figure 4). Although the chances of endorsing this item was monotonically ascendant, changes in the latent trait were just barely related to the probability of correctly endorsing it. Thus, the probability of correctly answering this item for anyone with $\theta = -2$ or -1 is about 0.35.

The most difficult (left side) and easiest items (right side) are presented in Figure 5. The participants should have an ability of 2.694 and -3.496 to have a .5 probability of correctly endorsing these items. The IRT results matched the simple proportion of correct responses: .934 for item 1 and .017 for item 27.

Results of invariance testing are presented in Table 2. The invariance analysis demonstrated that configural and weak invariance between males and females were both preserved. In the configural model, there were no constraints. In the weak invariance model, all factor loadings were constrained to be equal across groups. Metric invariance is established when a change in the model fit between the configural, and the metric invariance models is smaller than the tolerable change.

Table 1 Item descriptive, factor analysis, and IRT results.

Item	Descriptive			CFA		IRT	
	%	Item-Total <i>r</i>	α if deleted	λ	Std. λ	<i>a</i>	<i>b</i>
1	.934	.258	.878	1.000	.474	.847	-3.496
2	.876	.289	.878	1.135	.538	.777	-2.793
3	.922	.287	.878	1.291	.612	.930	-3.052
4	.916	.305	.877	1.295	.614	.935	-2.933
5	.485	.185	.882	.447	.212	.292	.214
6	.873	.293	.877	.955	.452	.794	-2.717
7	.770	.383	.876	1.061	.503	.921	-1.523
8	.698	.330	.877	.835	.396	.634	-1.426
9	.566	.314	.878	.800	.379	.625	-.455
10	.839	.362	.876	1.102	.523	.988	-1.975
11	.798	.438	.874	1.303	.618	1.239	-1.406
12	.715	.473	.873	1.309	.621	1.258	-.927
13	.824	.541	.872	1.718	.814	2.160	-1.186
14	.733	.510	.872	1.448	.686	1.552	-.897
15	.587	.458	.874	1.239	.587	1.235	-.343
16	.344	.386	.876	1.095	.519	1.068	.759
17	.534	.599	.870	1.700	.806	2.459	-.050
18	.479	.636	.869	1.830	.868	3.162	.119
19	.445	.676	.867	2.000	.948	5.343	.212
20	.440	.658	.868	2.050	.972	5.770	.224
21	.345	.628	.869	1.942	.920	4.566	.476
22	.266	.574	.871	1.821	.863	3.613	.715
23	.179	.471	.874	1.736	.823	2.983	1.051
24	.200	.519	.872	1.893	.897	4.053	.912
25	.119	.400	.875	1.701	.806	2.950	1.330
26	.076	.319	.877	1.507	.714	2.393	1.719
27	.017	.172	.879	1.289	.611	2.116	2.694
28	.033	.214	.879	1.246	.591	2.094	2.339

CFA: $\chi^2(350) = 5767.375, p < .001, CFI = .935, TLI = .930, RMSEA = .091$
 IRT: $M^2(350) = 2864, p < .001, CFI = .915, TLI = .909, RMSEA = .074$

Note. % = percentage of correct endorsing; Item-Total *r* = correlation between each item and total of the test; α if deleted = total Cronbach's alpha computed without each specific item; λ = factor loading; Std. λ = standardized factor loading; *a* = discrimination; *b* = difficulty.

Table 2 Results of invariance analysis

Invariance	Fixed	χ^2 (df)	CFI	Δ CFI	TLI	Δ TLI	RMSEA	Δ RMSEA
Configural	Equivalence	4434.5 (700)	.958		.955		.092	
Metric	Loadings	5089.3 (727)	.953	.006	.951	.004	.096	.004
Scalar	Intercept	4642.7 (726)	.956	.004	.955	.004	.092	.004

Following the evidence of the internal structure and reliability of the GMI, we tested differences between groups and computed normative data. This last step included percentiles, means, standard deviations, and IQ scores. Therefore, we addressed how demographic diversity might affect the research outcomes and the validity of the interpretation of the results. In addition, this study also provides evidence about the relationship between sociodemographic factors, such as age and educational background and intelligence. No outliers were observed in the distribution of the total scores ($G = 2.19$ and $G = 2.74$, both *p-values* = 1). The unrestricted model including a 3-factor interaction with sex, level of education, and age group was not identifiable due to aliased coefficients, meaning we had multicollinearity among predictors and thus interpreting the output is not advisable. A regression model that estimated the effect of education level, age group, and their interaction while controlling for the effect of gender was defined. This model was significant and accounted for approximately 16% of the

total variance ($R^2 = .16$, 95% confidence interval = [.12, .19]). The interaction was significant ($F(8,1259) = 4.89, p < .001, \eta_p^2 = .03$, 90% CI $\eta_p^2 = [.01, .04]$). The effects of level of education ($F(2,1259) = 19.89, p < .001, \eta_p^2 = .03$, 90% CI $\eta_p^2 = [.02, .05]$), age group ($F(4,1259) = 6.13, p < .001, \eta_p^2 = .02$, 90% CI $\eta_p^2 = [.01, .03]$), and sex ($F(1,1259) = 11.70, p < .001, \eta_p^2 = .01$, 90% CI $\eta_p^2 = [.00, .02]$) were significant.

Statistical results in which the interaction term in the model was significant indicated a moderation effect, as shown in Figure 6. Participants who completed only the elementary school level of education ($M = 7.36, SE = .905$) generally had lower results than participants who completed high school ($M = 12.05, SE = .58$) or who had an undergraduate level of education ($M = 13.94, SE = .508$). The *post hoc* analyses indicated that these results held only at ages between 15 and 25 years. At 36-45 years of age, participants who had an undergraduate level of education had higher results than participants with only an elementary school level, whereas the difference from participants with a high school education was not

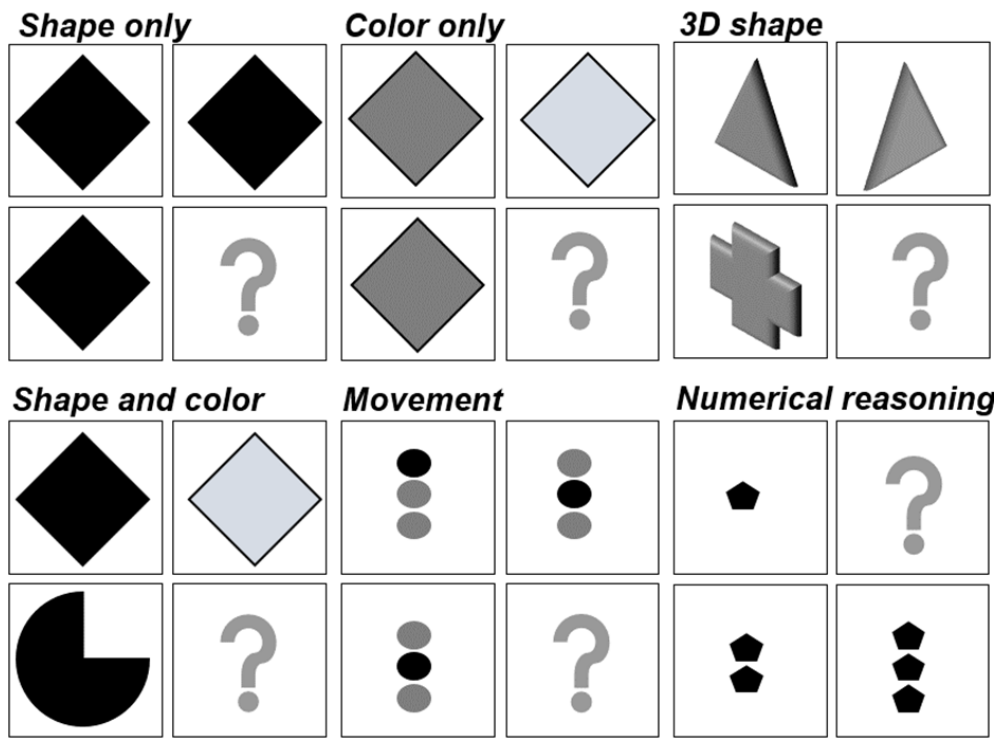


Figure 1
Example of items in the final version of the GMI (adapted due to copyrights).

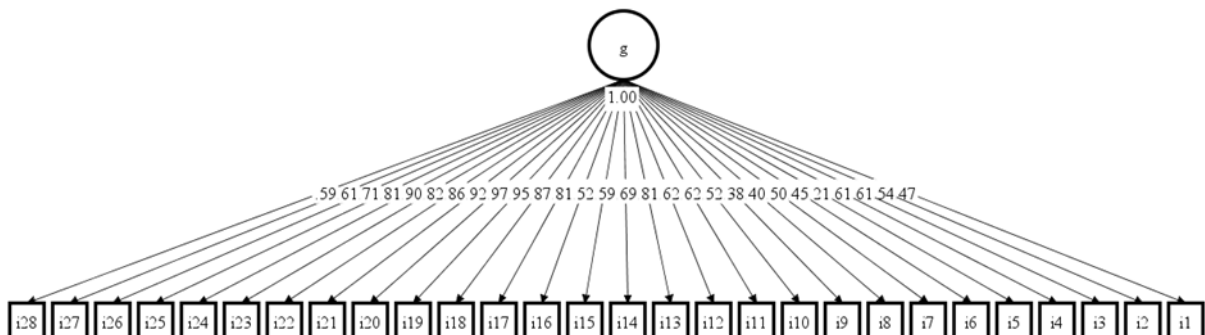


Figure 2
Factorial model of the GMI

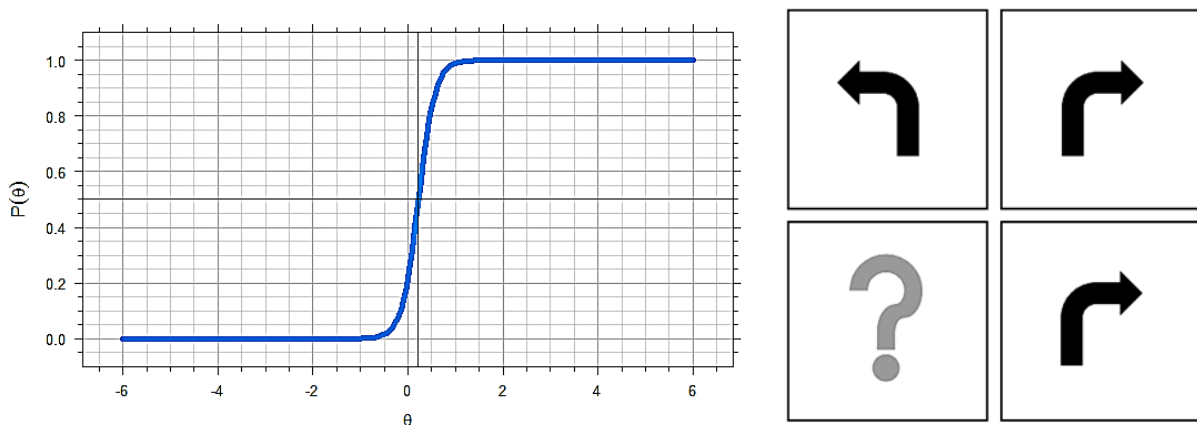


Figure 3
Item 20 curve analysis and stimulus (adapted due to copyrights)

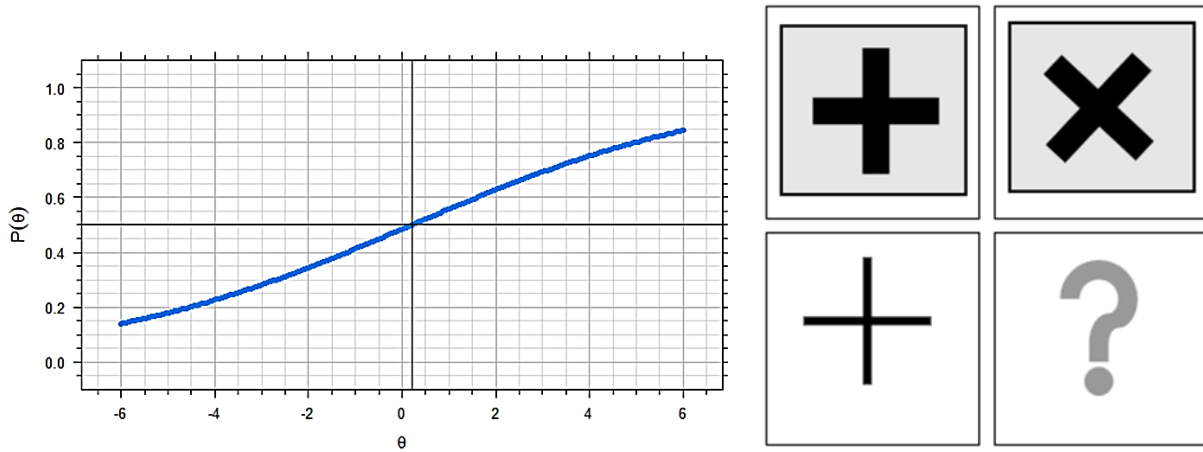


Figure 4
Item 5 curve analysis and stimulus (adapted due to copyrights)

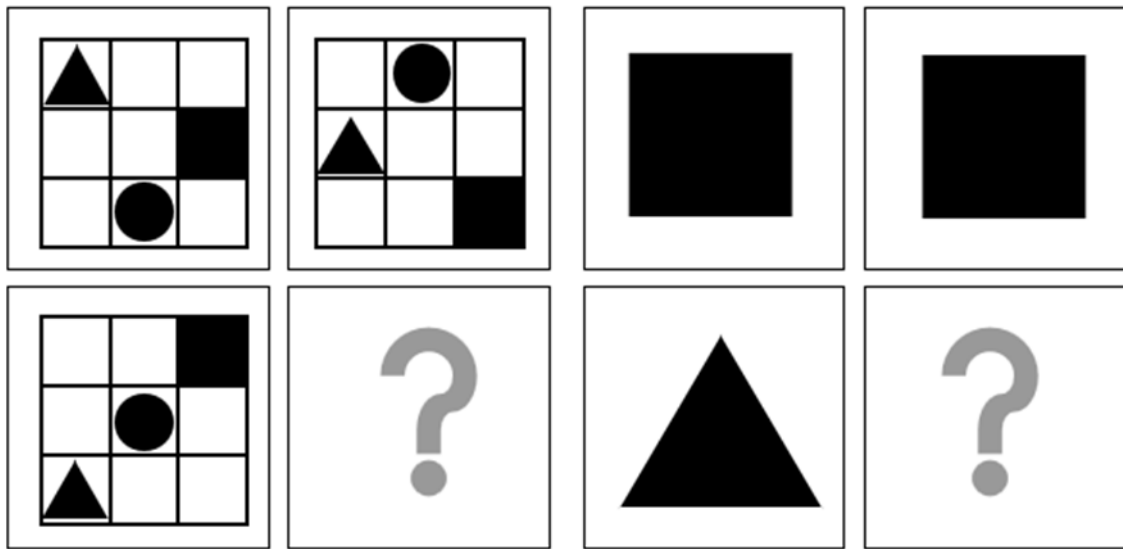


Figure 5
Item Curve Analysis (adapted due to copyrights)

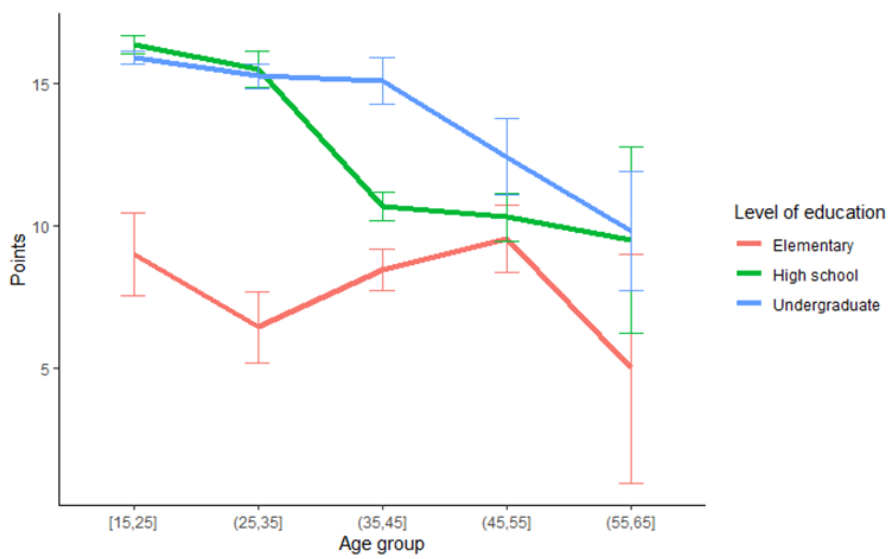


Figure 6
Average points by different age groups with different levels of education.

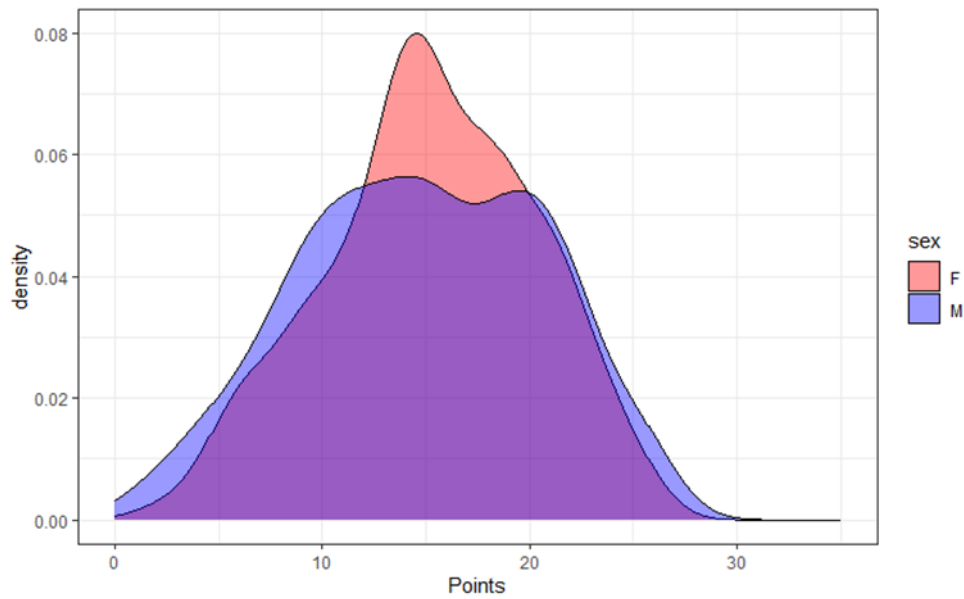


Figure 7
Overlap of male versus female distribution

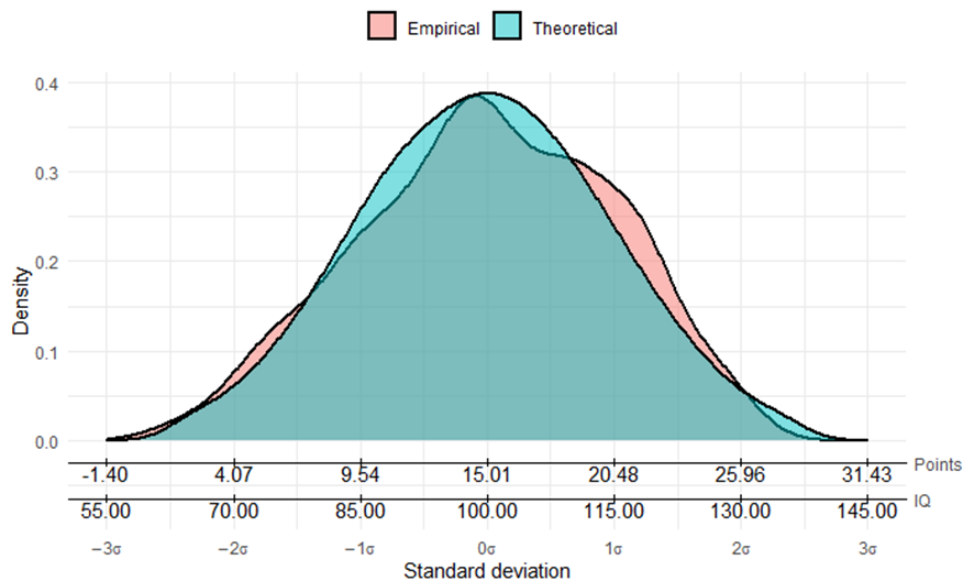


Figure 8
Empirical and theoretical distribution of the results.

DISCUSSION

The present study was conducted to evaluate the psychometric properties of a new matrix-based intelligence test, the GMI. In this initial study we investigated group differences and normative data based on means and IQ scores conversions to address the ongoing need for a brief non-verbal instrument in traffic psychology settings (Anuniação et al., 2021b). Psychologists working in this sector are required to assess the cognitive ability of people who apply for obtaining a national driver's license. Therefore, this study not only fills a gap in the Brazilian context but also enriches the psychometrics and intelligence assessment literature with a novel and culturally relevant assessment tool. Our results suggest

solid evidence for the connection between scores on the GMI and the g factor, a central measure of human intelligence. The unidimensional solution obtained in CFA and IRT suggests that the GMI effectively assesses cognitive abilities related to non-verbal general intelligence. This conclusion is supported by the internal consistency of the items, as demonstrated by high Cronbach's alpha and Guttman's lambda values. In previous studies that used Raven's Progressive Matrices (RPM) or RPM-based tests, some authors also achieved a one-factor structure (Lúcio et al., 2019; Waschl et al., 2016). Overall, the discrimination ability of all items was adequate, and an item in which a mirroring rotation ability was necessary to correctly respond led this parameter. Conversely, an item in which

the participant should rotate a geometric form was the least discriminative.

Testing whether a test is invariant across groups is required to determine whether the scores have the same meaning and interpretation across groups. Consequently, meeting this assumption is deemed required for fairness in testing (Walter et al., 2021). The invariance of a measure is an equivalence prerequisite for providing valid comparisons between different groups (Lacko et al., 2022). Current guidelines consider that invariance analysis is a crucial assumption for fairness in testing (Walter et al., 2021). Several types of method equivalence (e.g., method and item equivalence) are recommended in the literature. Our results provide initial evidence that males and females with equivalent trait levels would have the same probability of endorsing GMI-28 items. Although evidence in the literature is mixed, previous studies that applied the Raven Progressive Matrices (RPM) and Advanced Progressive Matrices (APM) found invariance between genders (Lúcio et al., 2019; Waschl et al., 2016). Lúcio et al. (2019) evaluated a sample of 582 Brazilian preschoolers and tested measurement invariance through multiple indicators multiple causes, in which sex and age were covariates. Their results were similar to ours but revealed that three items showed variance. Waschl et al. (2016) assessed 1297 individuals who completed the 12-item short form of the APM, 455 individuals who completed the full APM, and 362 individuals who completed a 15-item short form of the APM. Their results indicated that all three forms of the APM were unidimensional and measurement invariant across sex. However, other studies that examined the measurement invariance of cognitive ability tests across gender revealed that some tests only partially met the criteria for measurement invariance (Palejwala & Fine, 2015). For example, Immekus and Maller (2010) tested 2,000 adolescents using the Kaufman Adolescent and Adult Intelligence Scale and found intercept differences on the Memory for Block Designs, Famous Faces, Auditory Comprehension, and Logical Steps subtests, indicating low to moderately higher expected scores for males (Immekus & Maller, 2010). Keith et al. (2011) used the Differential Ability Scales, 2nd edition, with 2,600 children, 5-17 years old, and found that complete intercept invariance (i.e., scalar or strong) was not supported.

The analysis of trends and variability in normative data is crucial for understanding the distribution of intelligence within a population and interpreting IQ scores accurately. Variability in IQ scores can arise from various factors, including demographic characteristics, cultural differences, and methodological considerations. For example, age-related trends may influence IQ scores, with performance typically improving during childhood and adolescence before stabilizing or declining in adulthood (Schaie, 2013). Additionally, socio-economic status, educational opportunities, and environmental factors can contribute to variability in IQ scores, highlighting the importance of considering these factors when interpreting results (Nisbett et al., 2012).

Cultural factors may also play a significant role in shaping intelligence test performance, as individuals from different cultural backgrounds may approach tasks differently or have varying levels of exposure to the content covered in the tests (Sternberg, 2002). Moreover, methodological factors such as test administration, scoring procedures, and sample characteristics can introduce variability in normative data, influencing the interpretation of IQ scores (Reynolds & Keith, 2013). Understanding the sources of variability in normative data is essential for clinicians, educators, and researchers to accurately

interpret IQ scores and make informed decisions about individuals' cognitive abilities. By considering demographic characteristics, cultural background, and methodological factors, practitioners can ensure that IQ scores reflect true differences in cognitive functioning rather than artifacts of the testing process. Further exploring how the generated norms can be applied or adapted in various cultural contexts is essential for developing tests that are inclusive and fair across diverse populations. Cultural factors play a significant role in shaping individuals' cognitive processes, attitudes, and behaviors, influencing their performance on intelligence tests (Hambleton et al., 2005). Therefore, it is crucial to consider cultural differences when interpreting test scores and establishing norms to ensure the validity and fairness of assessments (Van de Vijver & Leung, 1997).

The findings regarding the impact of demographic variables such as age and education on test results shed light on the complex interplay between individual characteristics and cognitive performance. As evidenced by the significant effects observed in this study, age and education level emerged as key predictors of performance on intelligence tests. These results corroborate existing literature highlighting the influence of demographic factors on cognitive functioning (Salthouse, 2009). For instance, the observed decline in cognitive performance with advancing age aligns with the well-documented phenomenon of age-related cognitive decline, attributed to changes in brain structure and function over the lifespan (Nyberg et al., 2012). However, although several studies reported that advancing age is associated with a decrease in cognitive performance, novel studies that used latent growth models concluded that the rate of change in fluid intelligence has a positive, rather than negative slope (Kievit et al., 2018). We need to stress that with the data we are working on, we cannot make causal claims. As GMI is a brief measure that is completed quickly, the decline we found in older people may be a consequence of slowing reaction time instead of cognitive abilities.

Similarly, the significant differences in test scores across education levels underscore the role of education in shaping cognitive abilities, with higher levels of education generally associated with superior cognitive performance (Deary et al., 2010). The correlation between intelligence test scores and education level can be as high as .8 (Deary, 2012). Additionally, Ritchie and Tucker-Drob (2018) employed several regression models and concluded beneficial effects of education on cognitive ability of approximately 1 to 5 IQ points for each additional year of education.

The bell shape of the raw results indirectly indicates that the g factor could roughly follow a normal distribution, which also justified the procedure of scaling the raw distribution of scores into IQ scores to strengthen communication of the results. We stress that scores were not adjusted to match this distribution, a procedure that is traditionally performed with intelligence tests (Cramer, 2021). However, the shape of the data distribution was unexpected as we did not expect to find a normal distribution with the raw test data. As a typical procedure in intelligence tests, we were previously inclined toward the use of an algorithm to "fix" the empirical distribution and provide normalized standard scores with a mean of 100 and standard deviation of 15. However, our results were similar to Warne et al. (2013) and close to Rosseti et al. (2009) both of which achieved an empirical distribution that was roughly similar to a Gaussian function.

IMPLICATIONS FOR RESEARCH AND PRACTICE

The current study has several implications for research and practice. First, results of this initial pilot test of the GMI suggest that this brief non-verbal assessment is promising. However, follow-up evaluations of this measure are necessary to better establish the criterion-related and predictive validity of the GMI. Such evaluations should include widely accepted, valid measures of intelligence but should also evaluate how performance on the GMI predicts future performance driving and/or future performance on other societal tasks, such as responsible fire-arm ownership. Moreover, further evaluations of the GMI with larger, more diverse samples would contribute to the external validity of this measure. In the current study we carefully evaluated the performance of the participants according to the demographic characteristics of the sample and found invariance by gender, but we also found variations in performance by age and education level.

The normative data have important implications for intelligence assessment as they underscore the need for considering demographic factors when interpreting test results and making diagnostic or prognostic judgments. Failure to account for age-related changes or educational disparities may lead to misinterpretation of cognitive abilities and inaccurate assessment outcomes (Stern, 2012). Moreover, the observed interactions between demographic variables highlight the complex nature of cognitive development and underscore the importance of adopting a lifespan perspective in intelligence assessment (Schaie, 2013). Understanding how age and education influence test performance can inform the development of more tailored and culturally sensitive assessment protocols that account for individual differences and minimize bias (Weiss et al., 2010).

In the current study, we compared the performance of participants aged 15–64 years who had different education levels and provided normative data for the test. The significant difference in performance of males and females became nonsignificant after correcting the results. The empirical distribution of the data roughly followed a Gaussian function, which facilitated scaling the test results to IQ scores. By integrating insights from demographic research into assessment practices, clinicians and researchers can enhance the validity and utility of intelligence tests across different populations and contexts. Future research that continues to evaluate variations in performance based on these and other important individual characteristics (e.g., socio-economic status, broader geographic representation, etc.) are needed.

Additionally, the GMI can potentially have broader implications for intelligence assessment, enhancing the understanding of cognitive abilities and facilitating more accurate and equitable evaluations of individuals' cognitive functioning (Kaufman & Kaufman, 2004). The utilization of normative data derived from these findings holds crucial importance in practical applications. In clinical settings, the understanding of how demographic variables such as education level, age, and gender influence cognitive performance aids in accurate assessment and diagnosis of cognitive deficits (Matarazzo, 1990). For instance, clinicians can interpret test results more effectively by considering the expected performance range for individuals with different educational backgrounds and age groups. Because this test could be used as a cognitive screening tool, the use of the default neuropsychological metric of $\mu = 100$ and $\sigma = 15$ is useful for clinicians who work in this field.

Awareness of gender differences in cognitive performance informs tailored interventions and treatment plans (Halpern et al., 2007). Moreover, in contexts such as traffic assessment where cognitive abilities are essential for safe driving, these findings underscore the necessity of accounting for demographic variables in evaluating cognitive fitness for driving (Caird et al., 2014). By incorporating normative data derived from diverse demographic groups, assessments can be more equitable and reliable, leading to better-informed decisions regarding individuals' cognitive abilities and their implications for various contexts, thus promoting safety and fairness in assessment practices. However, it's essential to approach such interpretations cautiously, considering the dynamic nature of cognitive abilities and the potential biases inherent in standardized testing (McGrew, 2009). Finally, we believe the test presented in this study should be used as a tool to help clinicians improve people's lives. This could be done by offering them a better map of their psychological profile or adequate support when needed. This test should not be used to nurture or reinforce an outdated, prejudicial perspective about intelligence and achievement.

It is also important to consider the effects of technology on the application of intelligence instruments and test production. Discussing the role of technology and innovation in the future of intelligence research creates exciting possibilities for advancing our understanding of cognitive abilities. With the rapid advancements in technology, particularly in the field of artificial intelligence (AI), there is considerable potential for developing new tools and methodologies to measure and understand intelligence more holistically. AI algorithms can analyze vast amounts of data quickly and efficiently, allowing researchers to identify patterns and relationships that may not be apparent through traditional methods (Hutter et al., 2015). Machine learning techniques, for example, can be applied to large-scale datasets to uncover complex interactions between genetic, environmental, and cognitive factors that influence intelligence (Plomin & Deary, 2015). Moreover, emerging technologies such as virtual reality (VR) and augmented reality (AR) offer innovative ways to assess cognitive abilities in immersive and ecologically valid environments (Parsons & Rizzo, 2008). VR-based assessments can simulate real-world scenarios and challenges, providing insights into problem-solving skills, spatial reasoning, and decision-making abilities (Witmer & Singer, 1998). Additionally, wearable devices and mobile applications enable continuous monitoring of cognitive performance in daily life, offering valuable data for longitudinal studies and personalized interventions. However, it is essential to consider the ethical implications and potential biases associated with the use of technology in intelligence research. Issues such as data privacy, algorithmic transparency, and equitable access to technology need to be addressed to ensure that innovations in intelligence assessment benefit all individuals (Floridi et al., 2018).

In the realm of employment, advancements in intelligence research have implications for workforce training and recruitment practices. Employers increasingly recognize the value of cognitive diversity and seek to leverage individuals' unique strengths and talents (Schmidt & Hunter, 2004). By embracing a broader conception of intelligence that encompasses diverse skills and aptitudes, organizations can foster innovation and creativity in the workplace, driving economic growth and competitiveness (Sternberg, 2018b). Furthermore, intelligence research informs social policies aimed at reducing inequalities and

enhancing social mobility. By identifying factors that contribute to cognitive development and academic achievement, policymakers can implement targeted interventions to support vulnerable populations and break the cycle of poverty (Heckman, 2006). Investments in early childhood education, access to healthcare, and community development programs can promote cognitive development and improve life outcomes for individuals across the socio-economic spectrum (Noble et al., 2015).

Limitations

In addition to implications, there are several limitations of the present study that should be considered. First, our results were obtained in specific settings. Caution should be taken when extrapolating these findings to other groups. As stated in the introduction, many intelligence tests in Brazil are often normed on diverse populations, and therefore we recommend using these findings only for participants with similar characteristics to the ones presented. Second, the empirical distribution of the scores could be a consequence of the format of the test itself. During the developmental phase of the stimuli, steps were taken to avoid overly difficult or easy items. In a scenario where the goal of the test is to identify clinical or gifted participants, the empirical distribution of the results could be different. Third, the relationship between level of education and age group can also reflect the 6-minute administration time. We are not sure if the decrease in the score in older adults is, in reality, a consequence of this time constraint. Similarly, we could not formally test invariance for education level, which can distort the results. Finally, we could not provide sensitivity and specificity analyses. These last two limitations will be implemented in future studies.

CONCLUSION

Despite the aforementioned limitations, intelligence testing is highly valued and widely used in Brazil. Continuing to explore new approaches for conducting such testing efficiently and effectively will remain important as performance on these tests are used for gaining access to basic social rights, such as driving privileges. Therefore, continuing to evaluate how brief, culturally appropriate measures such as the GMI will remain important within Brazil. Such efforts should continue to explore the potential strengths of language-free assessments but should also take care to investigate how other potential factors such as gender and age may affect interpretations of assessment results.

ACKNOWLEDGMENTS

The authors have no support to report.

DECLARATION

i) Ethics approval and consent to participate

The research was previously approved by an ethics committee by Plataforma Brasil. All participants were presented with an Informed Consent Form and their doubts about the research procedures were clarified.

ii) Consent for publication

All authors agree to submit the article to the journal and grant copyright.

iii) Availability of Data and Material (ADM)

Codes that were used during the data analysis are freely available to ensure reproducibility of the results. All analyses were performed in R, with external packages. We provided a ready to run Rmarkdown on the command file folder at <https://osf.io/kvu42/>.

iv) Competing interests

The authors indicate that they have no conflicts of interest.

v) Funding

Not applicable.

vi) Authors contributions

All authors participated substantially in the preparation, review, and submission of the current manuscript.

REFERENCES

- Anderson, J. R. (1990). *Cognitive psychology and its implications* (3rd ed.). W H Freeman; Times Books; Henry Holt & Co.
- Alves, I. C. B., & Oliveira, M. S. (2017). Avaliação psicológica para motoristas profissionais: Fatores psicossociais e o teste psicológico como instrumento de intervenção [Psychological assessment for professional drivers: Psychosocial factors and psychological testing as an intervention tool]. *Psicologia: Teoria e Prática*, *19*(1), 139-151.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall/Pearson Education.
- Anstey, K. J., Windsor, T. D., Luszcz, M. A., & Andrews, G. R. (2006). Predicting Driving Cessation over 5 Years in Older Adults: Psychological Well-Being and Cognitive Competence Are Stronger Predictors than Physical Health. *Journal of the American Geriatrics Society*, *54*(1), 121-126. <https://doi.org/10.1111/j.1532-5415.2005.00471.x>
- Anstey, K. J., Wood, J., Caldwell, H., Kerr, G., & Lord, S. R. (2009). Comparison of Self-Reported Crashes, State Crash Records and an On-Road Driving Assessment in a Population-Based Sample of Drivers Aged 69-95 Years. *Traffic Injury Prevention*, *10*(1), 84-90. <https://doi.org/10.1080/15389580802486399>
- Anunciação, L., Portugal, A. C., & Landeira-Fernandez, J. (2021). Neuropsychological assessment: Statistical aspects of the relationship between percentile and classification. *Interação Em Psicologia*, *25*(3). <https://doi.org/10.5380/riep.v25i3.71375>
- Anunciação, L., Portugal, A., Rabelo, I., & Landeira-Fernandez, J. (2021). Non-verbal intelligence outperforms selective attention in a visual short-term memory test. *Psicologia: Reflexão e Crítica*, *34*(1), 35. <https://doi.org/10.1186/s41155-021-00200-0>
- Baker, F. B., & Kim, S. H. (2017). *The basics of item response theory using R*. Springer.
- Bünger, A., Grieder, S., Schweizer, F., & Grob, A. (2021). The comparability of intelligence test results: Group- and individual-level comparisons of seven intelligence

- tests. *Journal of School Psychology*, *88*, 101–117. <https://doi.org/10.1016/j.jsp.2021.09.002>
- Burns, E. (1988). Normalized standard score spreadsheet and norm table generator. *Journal of School Psychology*. [https://doi.org/10.1016/0022-4405\(88\)90038-6](https://doi.org/10.1016/0022-4405(88)90038-6)
- Caird, J. K., Johnston, K. A., Willness, C. R., Asbridge, M., & Steel, P. (2014). A meta-analysis of the effects of texting on driving. *Accident Analysis & Prevention*, *71*, 311–318. <https://doi.org/10.1016/j.aap.2014.06.005>
- Carroll, J. B. (2003). The Higher-stratum Structure of Cognitive Abilities. *The Scientific Study of General Intelligence* pp. 5–21. Elsevier. <https://doi.org/10.1016/B978-008043793-4/50036-2>
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. <https://doi.org/10.1037/h0046743>
- Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, *48*(6). <https://doi.org/10.18637/jss.v048.i06>
- 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>
- Ciotta, S., Cruz, R. M., & Dagostin, C. G.. (2021). Avaliação psicológica de condutores no trânsito no Brasil: marcos históricos e desafios atuais. *Revista Plural*, *1*(3)
- 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>
- Contran. (2022). Resolução nº 927, de 28 de março de 2022
- Conway, A. R. A., Kovacs, K., Hao, H., Rosales, K. P., & Snijder, J.-P. (2021). Individual Differences in Attention and Intelligence: A United Cognitive/Psychometric Approach. *Journal of Intelligence*, *9*(3), 34. <https://doi.org/10.3390/jintelligence9030034>
- Cramer, Y. (2021). Distribution across Age of Raw and Scaled Intelligence Scores - the DARSIS model. *PsyArXiv (OSF Preprints)* <https://doi.org/10.31234/osf.io/en82j>
- Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes: The PASS theory of intelligence*. Allyn & Bacon.
- Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, *63*(1), 453–482. <https://doi.org/10.1146/annurev-psych-120710-100353>
- Detterman, D. (2014). Editor's note. *Intelligence*, *42*, 135. <https://doi.org/10.1016/j.intell.2013.11.001>
- Flanagan JE, Landa R, Bhat A, et al. (2012) Head lag in infants at risk for autism: a preliminary study. *American Journal of Occupational Therapy*, *66*(5): 577–585
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.
- Goleman, D. (1995). *Emotional Intelligence: Why it can Matter More than IQ*. New York: Bantam.
- Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79–132. [https://doi.org/10.1016/S0160-2896\(97\)90014-3](https://doi.org/10.1016/S0160-2896(97)90014-3)
- Gottfredson, L., & Saklofske, D. H. (2009). Intelligence: Foundations and issues in assessment. *Canadian Psychology / Psychologie Canadienne*, *50*(3), 183–195. <https://doi.org/10.1037/a0016641>
- Guilmette, T. J., Sweet, J. J., Hebben, N., Koltai, D., Mahone, E. M., Spiegler, B. J., Stucky, K., & Westerveld, M. (2020). American Academy of Clinical Neuropsychology consensus conference statement on uniform labeling of performance test scores. *The Clinical Neuropsychologist*, *34*(3), 437–453. <https://doi.org/10.1080/13854046.2020.1722244>
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest* *8*(1):1–51. <https://doi.org/10.1111/j.1529-1006.2007.00032.x>
- Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Lawrence Erlbaum Associates.
- Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, *312*(5782), 1900–1902. <https://doi.org/10.1126/science.1128898>
- Herrnstein, R. J., & Murray, C. (2010). *The bell curve: Intelligence and class structure in American life*. Simon and Schuster.
- Horn, J. L., & McArdle, J. J. (2007). Understanding human intelligence since Spearman. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 205–247). Lawrence Erlbaum Associates Publishers.
- Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). (2015). *Automatic machine learning: methods, systems, challenges*. Springer.
- Immekus, J. C., & Maller, S. J. (2010). Factor structure invariance of the kaufman adolescent and adult intelligence test across male and female samples.

- Educational and Psychological Measurement*, 70(1), 91–104. <https://doi.org/10.1177/0013164409344491>
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger Publishers; Greenwood Publishing Group.
- Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, 32(1), 95–107. [https://doi.org/10.1016/S0160-2896\(03\)00062-X](https://doi.org/10.1016/S0160-2896(03)00062-X)
- Kaufman, A. S., & Kaufman, N. L. (2004). Kaufman assessment battery for children: Second edition (KABC-II). Circle Pines, MN: American Guidance Service.
- Keith, T. Z., Reynolds, M. R., Roberts, L. G., Winter, A. L., & Austin, C. A. (2011). Sex differences in latent cognitive abilities ages 5 to 17: Evidence from the Differential Ability Scales–Second Edition. *Intelligence*, 39(5), 389–404. <https://doi.org/10.1016/j.intell.2011.06.008>
- Kievit, R. A., Fuhrmann, D., Borgeest, G. S., Simpson-Kent, I. L., & Henson, R. N. A. (2018). The neural determinants of age-related changes in fluid intelligence: A pre-registered, longitudinal analysis in UK Biobank. *Wellcome Open Research*, 3, 38. <https://doi.org/10.12688/wellcomeopenres.14241.2>
- Koch, M., Becker, N., Spinath, F. M., & Greiff, S. (2021). Assessing intelligence without intelligence tests. Future perspectives. *Intelligence*, 89, 101596. <https://doi.org/10.1016/j.intell.2021.101596>
- Lacko, D., čeněk, J., Točík, J., Avsec, A., Đorđević, V., Genc, A., Haka, F., šakotić-Kurbalija, J., Mohorić, T., Neziri, I., & Subotić, S. (2022). The necessity of testing measurement invariance in cross-cultural research: Potential bias in cross-cultural comparisons with individualism–Collectivism self-report scales. *Cross-Cultural Research*, 56(2–3), 228–267. <https://doi.org/10.1177/10693971211068971>
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne M. (2009). Social science.Computational social science. *Science*, 6;323(5915):721–3. <https://doi.org/10.1126/science.1167742>
- Le Carret, N., Lafont, S., Letenneur, L., Dartigues, J.-F., Mayo, W., & Fabrigoule, C. (2003). The effect of education on cognitive performances and its implication for the constitution of the cognitive reserve. *Developmental Neuropsychology*, 23(3), 317–337. https://doi.org/10.1207/S15326942DN2303_1
- Lúcio, P. S., Cogo-Moreira, H., Puglisi, M., Polanczyk, G. V., & Little, T. D. (2019). Psychometric investigation of the raven's colored progressive matrices test in a sample of preschool children. *Assessment*, 26(7), 1399–1408. <https://doi.org/10.1177/1073191117740205>
- Matarazzo, J. D. (1990). Psychological assessment versus psychological testing: Validation from Binet to the school, college, and adult intelligence tests. *American Psychologist*, 45(9), 999–1017. <https://doi.org/10.1037//0003-066x.45.9.999>
- Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human Abilities: Emotional Intelligence. *Annual Review of Psychology*, 59(1), 507–536. <https://doi.org/10.1146/annurev.psych.59.103006.093646>
- 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>
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77–101. <https://doi.org/10.1037/0003-066X.51.2.77>
- Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *Am Psychol*, 67(2):130–59. <https://doi.org/10.1037/a0026699>
- Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Memory aging and brain maintenance. *Trends in Cognitive Sciences*, 16(5), 292–305. <https://doi.org/10.1016/j.tics.2012.04.005>
- Noble, K. G., Houston, S. M., Brito, N. H., Bartsch, H., Kan, E., Kuperman, J. M., Akshoomoff, N., Amaral, D. G., Bloss, C. S., Libiger, O., Schork, N. J., Murray, S. S., Casey, B. J., Chang, L., Ernst, T. M., Frazier, J. A., Gruen, J. R., Kennedy, D. N., Van Zijl, P., & Mostofsky, S. (2015). Family income, parental education and brain structure in children and adolescents. *Nature Neuroscience*, 18(5), 773–778. <https://doi.org/10.1038/nn.3983>
- Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(3), 250–261. <https://doi.org/10.1016/j.jbtep.2007.07.007>
- Palejwala, M. H., & Fine, J. G. (2015). Gender differences in latent cognitive abilities in children aged 2 to 7. *Intelligence*, 48, 96–108. <https://doi.org/10.1016/j.intell.2014.11.004>
- Plomin, R., & Deary, I. (2015). Genetics and intelligence differences: Five special findings. *Molecular Psychiatry*, 20 (1), 98–108. <https://doi.org/10.1038/mp.2014.105>
- Redvaldsen, D. (2020). Do the Mega and Titan tests yield accurate results? An investigation into two experimental intelligence tests. *Psych*, 2(2), 97–113. <https://doi.org/10.3390/psych2020010>
- Reynolds, C. R., & Kamphaus, R. W. (2003). Handbook of psychological assessment (4th ed.). Wiley.
- Reynolds, C. R., & Keith, T. Z. (2013). The future of assessment in psychology: predicting, explaining, and preventing behavior problems. Routledge.

- Rindermann, H., Becker, D., & Coyle, T. R. (2020). Survey of expert opinion on intelligence: Intelligence research, experts' background, controversial issues, and the media. *Intelligence*, *78*, 101406. <https://doi.org/10.1016/j.intell.2019.101406>
- Ritchie, S. J., & Tucker-Drob, E. M. (2018). How much does education improve intelligence? A meta-analysis. *Psychological Science*, *29*(8), 1358–1369. <https://doi.org/10.1177/0956797618774253>
- 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>
- Rosseti, M. O., Rabelo, I. S. A., Leme, I. F. A. de S., Pacanaro, S. V., & Güntert, I. B. (2009). Evidências de validade das Matrizes Progressivas Avançadas de Raven em universitários. *Psico-USF*, *14*(2), 177–184. <https://doi.org/10.1590/S1413-82712009000200006>
- Sattler, J. M. (2008). *Assessment of children: Cognitive foundations* (5th ed.). Jerome M. Sattler, Publisher, Inc.
- Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiol Aging*, *30*(4):507-14. <https://doi.org/10.1016/j.neurobiolaging.2008.09.023>
- Schaie, K. W. (2013). *Developmental influences on adult intelligence: The Seattle Longitudinal Study* (2nd ed.). Oxford University Press.
- Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173. <https://doi.org/10.1037/0022-3514.86.1.162>
- Silva, F. H. V. C. , & Günther, H. (2009). Psicologia do trânsito no Brasil: de onde veio e para onde caminha?. *Temas em Psicologia*, *17*(1), 163-175.
- Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*. <https://doi.org/10.2307/1412107>
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge University Press.
- Sternberg, R. J., & Grigorenko, E. L. (2002). The theory of successful intelligence as a basis for instruction and assessment. ASCD.
- Sternberg, R. J. (2017). *Wisdom, intelligence, and creativity synthesized*. Cambridge University Press.
- Sternberg, R. J. (2018a). Theories of intelligence. *APA Handbook of Giftedness and Talent*, 145–161. American Psychological Association. <https://doi.org/10.1037/0000038-010>
- Sternberg, R. J. (2018b). *Applying intelligence to the workplace*. Routledge.
- Stern, Y. (2012). Cognitive reserve in aging and Alzheimer's disease. *Lancet Neurol*, *11*(11):1006-12. [https://doi.org/10.1016/s1474-4422\(12\)70191-6](https://doi.org/10.1016/s1474-4422(12)70191-6)
- Thurstone, L.L. (1938). *Primary mental abilities*. University of Chicago Press: Chicago.
- Tomlinson, C. A. (2017). *How to differentiate instruction in academically diverse classrooms*. ASCD.
- Van Der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. <https://doi.org/10.1037/0033-295X.113.4.842>
- Van de Vijver, F. J., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Sage Publications.
- Warne, R. T., & Burningham, C. (2019). Spearman's g found in 31 non-Western nations: Strong evidence that g is a universal phenomenon. *Psychological Bulletin*, *145*(3), 237–272. <https://doi.org/10.1037/bul0000184>
- Waterhouse, L. (2006). Inadequate Evidence for Multiple Intelligences, Mozart Effect, and Emotional Intelligence Theories. *Educational Psychologist*, *41*(4), 247–255. https://doi.org/10.1207/s15326985ep4104_5
- Walter, F., Daseking, M., & Pauls, F. (2021). Sex Differences in Intelligence in Children Aged 2:6–7:7: Analysis of the Factor Structure and Measurement Invariance of the German Wechsler Primary and Preschool Scale of Intelligence—Fourth Edition. *Journal of Psychoeducational Assessment*, *39*(4), 395–421. <https://doi.org/10.1177/0734282920981398>
- Warne, R. T., & Burningham, C. (2019). Spearman's g found in 31 non-Western nations: Strong evidence that g is a universal phenomenon. *Psychological Bulletin*, *145*(3), 237–272. <https://doi.org/10.1037/bul0000184>
- Warne, R. T., Godwin, L. R., & Smith, K. V. (2013). Are There More Gifted People Than Would Be Expected in a Normal Distribution? An Investigation of the Overabundance Hypothesis. *Journal of Advanced Academics*, *24*(4), 224–241. <https://doi.org/10.1177/1932202X13507969>
- Waschl, N. A., Nettelbeck, T., Jackson, S. A., & Burns, N. R. (2016). Dimensionality of the Raven's Advanced Progressive Matrices: Sex differences and visuospatial ability. *Personality and Individual Differences*, *100*, 157–166. <https://doi.org/10.1016/j.paid.2015.12.008>
- Waterhouse, L. (2006). Inadequate Evidence for Multiple Intelligences, Mozart Effect, and Emotional Intelligence Theories. *Educational Psychologist*, *41*(4), 247–255. https://doi.org/10.1207/s15326985ep4104_5
- Weiss, L. G., Saklofske, D. H., Prifitera, A., & Holdnack, J. A. (Eds.). (2010). *WISC-V assessment and interpretation: Scientist-practitioner perspectives*. Academic Press.
- Weinberg, R. A. (1989). Intelligence and IQ: Landmark issues and great debates. *American Psychologist*, *44*(2), 98–104. <https://doi.org/10.1037/0003-066X.44.2.98>
- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. <https://doi.org/10.1162/105474698565686>

This page has been intentionally left blank