Research Articles // Artículos de investigación

Notes and Editorial Information // Avisos e información editorial

ISSN 1577-7057

Volume 25, number 2 June 2025
Volumen 25, número 2 Junio 2025

ISSN: 1577-7057

IJP&PT

International Journal of Psychology & Psychological Therapy

IJP&PT 25 YEARS ANNIVERSARY 2001-2025

https://www.ijpsy.com

IJP&PT is included in the following indexing and documentation Centers:

University Grants Commission
Bahadur Shah Zafar Marg, New Delhi-110002
Website: www.ugc.ac.in

International Journal of Psychology & Psychological Therapy, 2025, 25, 2, 223-244

Psychometric Evaluation of the Relational Abilities Index in Mexican Population

Oscar Giovanni Balderas Trejo[*], **María Luisa Cepeda Islas, Hortensia Hickman Rodríguez**
Universidad Nacional Autónoma de México, México
Miguel Rodríguez Valverde
Universidad de Jaén, España

Abstract

This study analyzes the psychometric properties of the Spanish adaptation of the *Relational Abilities Index* (RAI) among university students in Mexico. The study employed a version of the RAI consisting of 128 items distributed across eight types of relational responses. The study evaluates its concurrent validity by calculating the relationship between relational abilities and intelligence, measured through the *Wechsler Adult Intelligence Scale* IV (WAIS-IV) and the *G Factor* test. The sample consisted of 40 university students, selected through non-probabilistic sampling. Results indicated moderate correlations between RAI scores and WAIS-IV subscale scores, particularly with visuospatial and abstract reasoning skills. Analyses of internal consistency of the RAI yielded a high Cronbach's alpha coefficient, although split-half reliability was low, suggesting a need to review item homogeneity. This study provides preliminary evidence on the validity and reliability of the RAI in a Mexican context, pointing out the importance of adjusting item difficulty and discrimination to enhance psychometric accuracy. Implications for the assessment and training of relational skills in educational and clinical settings are discussed.

Key words: relational skills, Relational Abilities Index, psychometric assessment, cultural adaptation, college students.

How to cite this paper: Balderas-Trejo OG, Cepeda-Islas L, Hickman-Rodríguez H, & Rodríguez-Valverde M (2025). Psychometric Evaluation of the Relational Abilities Index in Mexican Population. *International Journal of Psychology & Psychological Therapy, 25,* 2, 223-244.

Novelty and Significance

What is already known about the topic?

- RAI has shown moderate to strong correlations with intelligence tests like WAIS-III.
- Previous studies indicated challenges in split-half reliability and test consistency.

What this paper adds?

- The study evaluates the psychometric properties of the Spanish adaptation of the Relational Abilities Index in Mexican university students.
- This study demonstrates the RAI's high internal consistency but low split-half reliability
- Novel correlations between RAI scores and visuospatial and abstract reasoning skills were found.
- Highlights the need to adjust item difficulty and discrimination for enhanced psychometric accuracy.

Learning is a central topic in psychological research that has been addressed from different perspectives. De Houwer, Barnes-Holmes, & Moors (2013) propose to define learning as changes in the behavior of an organism that result from regularities in the environment; this definition leads us to the recognition of a functional relationship between behavior and environment. Relational behavior can be established as a product of learning. It can be defined as the ability to derive the relationship between two or more stimuli based on their context and without a direct history of reinforcement (Fryling, Rehfeldt, Tarbox, & Hayes, 2020). Relational Frame Theory (RFT) offers a

[*] *Correspondence*: Oscar Giovanni Balderas Trejo, Universidad Nacional Autónoma de México, de los Barrios Ave., Los Reyes Iztacala Tlalnepantla, Estado de México, CP 54090, México. E-mail: gio.balderas.trejo@gmail.com. *Acknowledgments*:This research was carried out with the support of SECIHTI granted to CVU 1147183, and the support of the PAPIIT-UNAM program IN300723.

broad and generalized approach to the study of relational behavior, conceptualizing verbal behavior and cognition as arbitrarily applicable relational responding (Mulhern, 2022). Hughes & Barnes-Holmes (2015) described relational behavior as responding to at least one stimulus in terms of another based on specific contextual cues that determine what relationship they have with each other.

In addition to focusing on the analysis of the characteristics of relational responses, it is essential to consider the role of Intelligence Quotient (IQ) as an indicator correlated with learning. IQ is assessed by a variety of standardized tests, which examine skills such as logical reasoning, memory, attention, and speed of processing, among others (Schneider & Newman, 2015). Among these tests, the *Wechsler Adult Intelligence Scale* (WAIS) stands out. This standardized test is used to measure intelligence and cognitive skills in adults and adolescents over 16 years of age (Úbeda, Fuentes, & Dasí, 2016).

The connection between IQ and relational behavior is relevant since individuals with higher IQ may show a greater aptitude for understanding and applying stimulus relations, which could drive the acquisition and application of more sophisticated cognitive and social skills. In contrast, individuals with lower IQ may have problems in establishing and manipulating relational responses, which could impact their performance in tasks that demand advanced cognitive skills. Also, it should be considered that IQ is not immutable and can be affected by factors such as education, environment, and interventions that focus on improving relational skills (Cassidy, Roche. & Hayes, 2011). Therefore, investigating relational responding in conjunction with IQ provides a broad approach to understanding and enhancing cognitive skills and performance in different aspects of a person's life.

Cassidy *et alii* (2011) conducted the first systematized test of relational skills in instrument, which was administered through the public website *RaiseYourIQ.com*. The assessment presented participants with 55 syllogistic relational items, which required approximately 14 minutes to complete. It consisted of 29 same/opposite tasks and 26 greater than/or less than tasks that progressed in difficulty across trials. A total of 248 stimuli, consisting of three-letter nonsense words (e.g., BEF, DIL, FAS), were used, with no stimuli repeated. Participants were informed beforehand that they would be asked questions about the relationships between these nonsense words, and that they could arrive at the answer by reading the statements carefully. The questions remained on screen until the user responded, however, a time limit of 30s per item was applied Failure to respond within this interval was considered an incorrect response and the participant moved on to the next item. The total number of correct responses produced within the 55 items was considered as the metric of relational skill mastery.

Subsequently, Cassidy, Roche, Colbert, Stewart, & Grey (2016) replicated the previous research, in their research the authors describe the presented sequence of the items where it can be observed that the items increase in complexity as the test progresses. For each test stage there are several possible relational tasks within the definition parameters. These are always randomly selected (within the parameters) by the computer software, therefore, no two participants are assigned the same tasks, furthermore, none of the stimuli are presented twice for a single participant, so each and every task presented on the screen is unique.

Colbert, Dobutowitsch, Roche, & Brophy (2017) conducted a validation of this version of the RAI (Relational Abilities Index) by correlating it with different intelligence tests, most notably with the WAIS-III. Each participant was administered the 13 core subtests of the WAIS-III, allowing the calculation of seven IQ indices and subindices.

These subtests included picture completion, vocabulary, digit and symbol coding, similarity, block design, arithmetics, matrix reasoning, digit span, information, picture layout, comprehension, symbol search, and letter and number sequencing.

RAI scores correlated moderately to strongly with full scale, verbal and performance IQ scores, as well as with the four IQ subindices (verbal comprehension, working memory, perceptual organization, and processing speed). With respect to the verbal subtests, strong correlations were found between RAI scores and arithmetic, comprehension, vocabulary and information, moderate correlations found for similarity, digit span, and letter and number sequence. As for the performance subtests, RAI scores showed a moderate significant correlation with block layout, matrix reasoning, and symbol search scores, but not with picture completion, digit and symbol coding, or picture arrangement scores. The authors highlighted the utility of the RAI in assessing both conceptual and functional intelligence skills, particularly in situations where an estimation of intellectual ability rather than a precise and direct measurement is required (Colbert *et alii*, 2017).

Despite the significance of these findings, the RAI was compared with tests that mainly evaluate crystallized intelligence, which encompasses knowledge acquired by an individual and their ability to utilize that knowledge, such as semantic knowledge, vocabulary, and the ability to comprehend and utilize language, which are typically acquired from a culture and heavily influenced by educational and social experiences. However, the Cattell-Horn-Carroll theory of intelligence includes a second factor, Fluid Intelligence, understood as the ability to reason and think abstractly, without relying on previously learned knowledge and experiences (Flanagan, 2014).

It is essential to evaluate the RAI against tests that assess intelligence through abstract tasks that are not dependent on prior knowledge, such as the G Factor test. This test aims to measure general intelligence based on performance on a series of relation Matrix Reasoning. It has been extensively discussed within the context of the Cattell-Horn-Carroll theory of cognitive abilities, which identifies several underlying cognitive abilities that contribute to general intelligence.

Colbert, Malone, Barrett, & Roche (2020) performed a validation of this version of the RAI by correlating the test results with the WAIS-III. Ninety-seven individuals between 18 and 45 years of age participated. The results of the correlational analysis revealed the presence of a significant relationship between the scores of RAI, full-scale IQ, verbal and WAIS-III scale. There is considerable overlap with the results of Colbert *et alii* (2017). Although the pattern of significant relationships was similar, the magnitude of correlations varied between these studies by Colbert and colleagues, with slightly lower correlations for Colbert et alii (2017). This latter study concluded that the addition of more relational frames in the assessment was not beneficial in improving the predictive utility of the RAI, as this inclusion increased the instrument's variance. A relevant finding was the lack of correlation between analogical skills and IQ. This may be accounted for by the limited number of items on analogy relations. In spite of it, the literature suggests that, due to the advanced level of complexity inherent to analogical reasoning, for example, its proficiency levels should predict IQ especially for high ability individuals; this may also be attributed to the low correlation of analogy test scores compared to scores obtained in tests of lower complexity.

Ruiz, Cepeda Islas, Hernández Miranda, Hickman Rodríguez, & Balderas Trejo (2022) examined how variations in trial length affect participants' performance on the RAI employed by Cassidy *et alii* (2016) and Colbert *et alii* (2017). Sixty-two undergraduates participated, randomly assigned to groups of 30 seconds or 20 seconds

per trial. The results showed that reducing trial time decreased the number of correct responses, especially in lower ability participants, indicating an increase in trial difficulty; however, in higher ability participants the same effect was not found. The study concluded that shorter trial durations could improve skill discrimination for medium and high ability participants, proposing an adjustment of trial duration to 15 seconds to improve discrimination indication in high-scoring participants.

Cummins, Nevejans, Colbert, & De Houwer (2023) noted that conceiving of relational responding as a form of generalized operant behavior means that the ability to establish relationships can be modified and improved through relational training of types of relations other than the tested skill, and by extension, this could improve intellectual skills; this was demonstrated in a recent meta-analysis which found a moderate overall impact of relational training on non-verbal IQ, although this analysis also noted that there are few comparable studies that have investigated this and a relatively high risk of bias in those existing studies (May, Tyndall, McTiernan, Roderique-Davies, & McLoughlin, 2022). They employed a version of the RAI that included 128 items divided into eight relationship types (opposite, different, containment, temporal, quantity, mathematical, analogy, and deictic), the tests of each relationship type progressively increased in difficulty along a series of dimensions, indicating that this is the first version of the RAI that assesses complex relationships considering the progressive increase in complexity. They found this version to be a reliable measure of relational responding, showing high test-retest reliability and split-half reliability, but with substantial variation across different subscales.

Subsequently, Cummins (2023) explored the importance of measuring relational responsiveness and the psychometric properties of the RAI (Cummins *et alii*, 2023). The sample included 264 typically developing adult participants between the ages of 18 and 40. The main results indicate that, although the RAI showed promise on overall measures, the reliability of the subscales varied, presenting challenges in terms of consistency. Split-half reliability and 1-week test-retest reliability was generally poor across the different subscales. Specifically, split-half reliability ranged from 0.68 to 0.96, and test-retest reliability was consistently low, ranging from 0.55 to 0.59. In addition, the analyses showed patterns of response error that did not follow the RFT predictions regarding difficulty of the relational types. This study suggested that, although the RAI is promising for measuring relational responding, improvements in the accuracy and consistency of the subscales are crucial for future research. This research proposes that, while the RAI shows promise in assessing relational responding, enhancing the precision and reliability of its subscales is imperative for future studies. The author underscored the necessity of enhancing the test's discrimination index and suggested that increasing the number of trials in the RAI could enhance its precision; nonetheless, this might entail an unwarranted proliferation of items, thus an alternative approach for enhancement would be to elevate the items' difficulty level.

While Cummins' version of the RAI (Cummins, 2023) represents an approximation for the assessment of complex relational skills, it is necessary to evaluate the psychometric properties of the instrument; in addition to this, it is proposed to generate an adaptation of the test to Spanish, so it is necessary to evaluate this new version of the RAI. This study aims to examine the psychometric properties of a Spanish adaptation of the RAI.

Method

Participants

An study was proposed with a sample of at least 21 undergraduate students from a public university in Mexico City. To estimate the sample size, a statistical power analysis was performed considering a value of α= 0.05, a power of 1-β= .80 and an effect size of 0.5 (Faul, Erdfelder, Buchner, & Lang, 2009), based on the values reported by May *et alii* (2022). Non-probabilistic sampling was used to select participants, establishing as exclusion criteria: having taken an IQ test in the past 6 months, having a diagnosed neurological or psychiatric condition that could influence IQ, having a visual or hearing impairment that interferes with the administration of the tests, and being under the influence of medications or substances that affect cognitive abilities during the evaluations. Subsequent discovery that the participant did not meet one or more of the exclusion criteria, as well as complete nonresponse to one or more of the assessments, was considered an elimination criterion.

This study was approved by the Ethics Committee of the Graduate Program in Psychology at the *Universidad Nacional Autónoma de México* (UNAM). All participants provided informed consent prior to their participation, ensuring their understanding of the procedures and their right to withdraw at any time without adverse consequences.

Two groups from the psychology program were invited to participate, obtaining an initial sample of 49 participants. Nine were eliminated for not completing the evaluations, resulting in a final sample of 40 participants. During the evaluation of the sensitivity achieved, an effect size of 0.37 was detected, which is below the parameters reported by May *et alii* (2022). There is a significant probability of detecting an effect, even if it is smaller than anticipated. This reduces the probability of making a Type II error, where a true effect is failed to be detected.

Instruments

Relational Abilities Index (RAI; Cummins *et alii*, 2023). The RAI is a computerized test designed to assess relational reasoning skills across eight types of relations: opposition, difference, containment, temporal, quantity, mathematical, analogy, and deictic. It includes 128 items with nonsense three-letter words, progressively increasing in difficulty. Each relational type is represented by 16 items. Items are time-limited to 30 seconds. An example item would present statements like "BEF is more than DIL. DIL is more than FAS" and ask the participant to derive the correct relationship. This version was adapted to Spanish following ITC Guidelines for Translating and Adapting Tests. Internal consistency in the current sample was high (α= .88). The design of the RAI aims to measure participants' capacity to establish relations between stimuli and concepts, based on specific contextual cues, without a direct history of reinforcement. The items use three-letter nonsense words to avoid familiarity biases and ensure that responses are based on relational capability rather than prior knowledge. This approach seeks to provide a reliable and valid measure of relational skills, considered fundamental for various cognitive and social functions. Additionally, the items are presented in a format that requires participants to make quick assessments and connections, with a response time limit for each item, adding a speed component to the assessment of relational ability. For further details about the types and progression of the items, refer to Appendix A and B.

Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV; Úbeda *et alii*, 2016; Sánchez Escobedo, 2015). The WAIS-IV is a standardized test used to assess general cognitive ability in individuals aged 16 and older. The present study evaluated four subtests reported by Colbert, Tyndall, Roche, & Cassidy (2018): Block Design, Matrix

Reasoning, Similarities, and Vocabulary. These subtests cover visual-spatial reasoning and verbal comprehension. The WAIS-IV has been validated and standardized for use in the Mexican population.

Culture Fair Intelligence Test (G Factor; Cattell *et alii*, 2017). This non-verbal intelligence test is based on the Cattell-Horn-Carroll theory of cognitive abilities. It measures fluid intelligence through abstract reasoning tasks using matrices. The version employed in this study was standardized for the Mexican population (González Velázquez, Aragón Borja, & Silva Rodríguez, 2000). The test was administered in a group setting using its Spanish adaptation.

Procedure

For the application of the tests, we began with the application of the WAIS-IV test to individual participants at the *Unidad de Investigación Interdisciplinaria en Ciencias de la Salud y Educación* of the FES Iztacala, UNAM. Once the evaluation with the WAIS-IV test was concluded, the evaluation of the RAI and G Factor tests was carried out as a group in digital format using computers in laboratories of the same university.

Data Analysis

The results of the RAI were analyzed in terms of frequency distribution, time of completion, and scores by type of relationship. The internal consistency of the test was estimated through the correlation between the scores of each type of relationship and the overall score, indicating an assessment of construct validity to ensure that all items measure the same underlying theoretical construct. The consistency per item was further validated through Cronbach's alpha coefficient.

To determine discriminant validity, the Kruskal-Wallis test was performed to evaluate whether there are statistically significant differences in the median scores between the different types of relationships. After calculating the Kruskal-Wallis test, the post hoc tests were performed to evaluate which specific relationship types differ from each other, using Dunn test and Bonferroni method to adjust for multiple comparisons and control for Type I error rate. This helps to assess whether the RAI can distinguish between different types of relational reasoning.

Additionally, convergent validity was explored by evaluating the frequency distribution of the WAIS-IV and G Factor tests. Correlations were established between the scores of the RAI, the WAIS-IV tests, and G Factor, as well as between the subscales of the WAIS-IV and the types of relationships evaluated in the RAI. This analysis checks whether the RAI measures aspects of intelligence that are theoretically expected to be related to established intelligence tests.

The analyses were performed using the R studio application, using the tidyverse, psych, dunn.test, CTT, and corrplot libraries; the plots were generated using the *ggplot2 library*.

RESULTS

The estimation of RAI scores shows that the *M*=85.92, *Med*=87, and a *SD*= 14.31with values ranging from 60 to 116 correct answers out of a possible 128, the standard error of the mean was estimated at 2.26. They present a skewness of -0.21, indicating a slight skewness to the left, kurtosis of -0.81, suggesting that the distribution is slightly platykurtic (see Figure 1). However, when assessing the normality of the

distribution by Shapiro-Wilk test we obtained W= 0.96 p= .21 which suggests a normal distribution of the data, this can be visually verified in Figure 2.

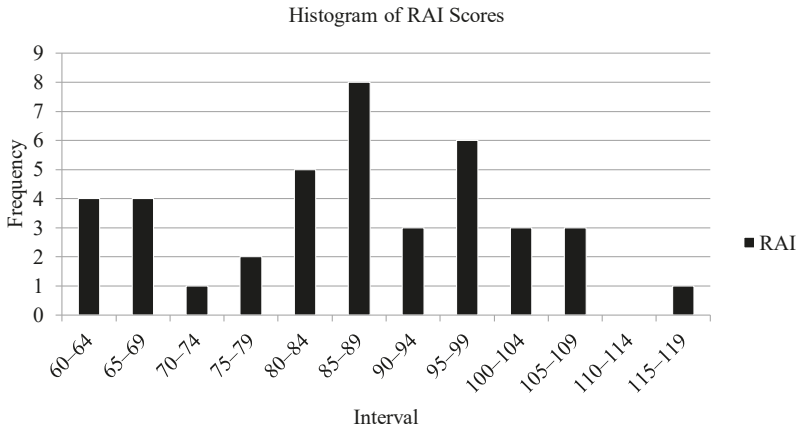

Figure 1. Histogram of the scores obtained in the RAI.

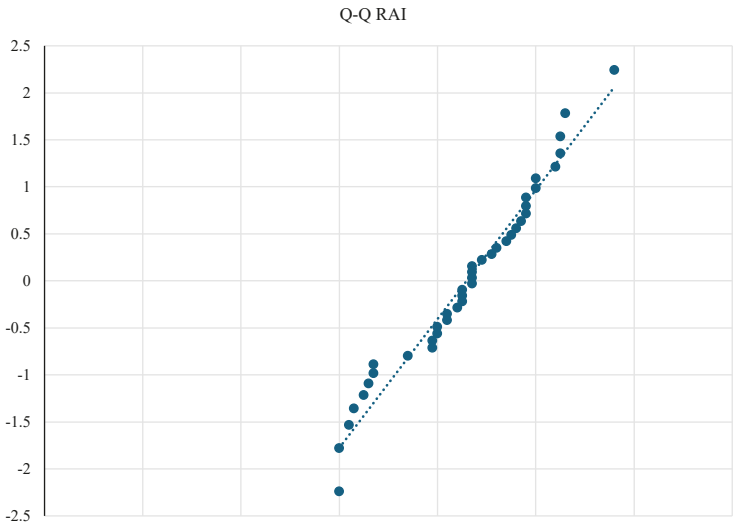

Figure 2. Q-Q plot of the scores obtained in the RAI.

Regarding the estimation of response times per item of the RAI, an average response time of 15.4 seconds per item with a deviation of 2.86 was presented, the response times varied between 9.15 and 21.1 seconds with a *Standard Error* of the Mean of 0.45 (see Figure 3). When evaluating the distribution of the times, a skewness of -0.001 and a kurtosis of -0.78 were found, indicating a better distribution with respect to a normal distribution compared to the distribution of the scores of the same instrument. To confirm this, the Shapiro-Wilk test was performed, and the results were obtained indicating that a normal distribution is indeed present (W= 0.98; p= .21).

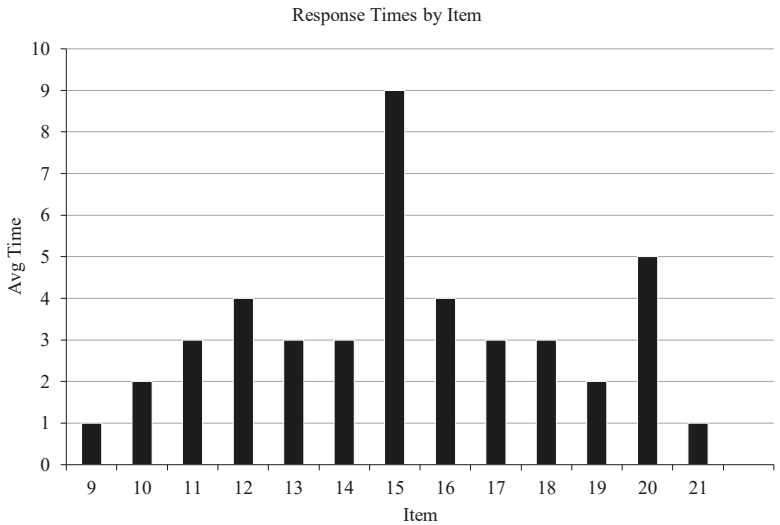

Figure 3. Histogram of response times by item in the RAI.

Table 1 shows the results obtained by type of relationship evaluated within the RAI. Opposition has a mean of 8.9 with a distribution slightly skewed to the left and flatter than a normal distribution, with a Shapiro-Wilk *p*-value suggesting a non-parametric distribution (p= .04). Difference has a higher mean of 13.65 and greater skewness to the left, along with a strong negative kurtosis, the Shapiro-Wilk test rejects normality. For the Quantile ratios a mean of 12.2 was estimated, it shows a slight skewness to the left and a relatively flat distribution, a distribution other than normal is assumed with a value of p= .02 Temporal has a mean of 10.45 and is the closest to a symmetrical distribution, it presents a Shapiro-Wilk *p*-value suggesting normality (p= .07). The Containment ratios have a mean of 12.45, show a skewness similar to that of Quantities and a kurtosis indicating a flat distribution; the Shapiro-Wilk test suggests a distribution other than normality (p= .01). Analogy-type relationships have a mean of 9.85 and show a slightly skewed distribution to the left and flat, with a Shapiro-Wilk test indicating non-normality (p= .01). For the Deictic relations it presents a mean of 9.93 and is the only one with a significant positive skewness, suggesting that the scores are skewed

Table 1. Descriptive statistics and normality tests by RAI relationship type.

Type of relationship	*M*	*SD*	*Med*	*min*	*max*	range	skew	kurtosis	*SE*	*W*	*p*
Opposition	8.9	3.2	9	3	14	11	-0.36	-0.96	0.50	0.94	.05
Difference	13.65	2.7	15	8	16	8	-0.77	-1.00	0.42	0.81	.00
Quantities	12.2	2.9	13	5	16	11	-0.57	-0.56	0.45	0.94	.02
Temporal	10.45	3.1	10	4	15	11	-0.11	-1.00	0.49	0.95	.07
Containment	12.45	2.9	12.5	6	16	10	-0.46	-0.81	0.46	0.92	.01
Analogy	9.85	2.0	10	5	13	8	-0.60	-0.52	0.32	0.92	.01
Deictic	9.92	2.5	10	6	16	10	0.65	-0.26	0.40	0.93	.01
Mathematical	8.5	2.2	9	4	14	10	0.20	-0.35	0.36	0.97	.48

Notes: *max*= maximum; *M*= mean; *Med*= median; *min*= minimum; *SD*= Standard Deviation; *SE*= Standard Error; skew= skewness; *W*= Shapiro-Wilk statistic.

towards the lower ones; the Shapiro-Wilk test suggests a distribution different from normality (p= .01). Finally, Mathematical has the lowest mean of 8.5, a slight skewness to the right, and a slightly flatter than normal distribution; however, its Shapiro-Wilk p-value indicates a normal distribution p= 0.48.

Relationships of the Difference type have the highest mean, showing a significant negative bias. On the other hand, Mathematical has the lowest mean, showing an inverse bias to that found in Difference, without being as marked. The variability remains between values of SD 2.2 to 3.1, with Temporal being the type of relationship with the highest standard deviation value and the highest kurtosis. Additionally, a high variability can be considered from the range between the minimum and maximum values, since six of the relationship types present ranges of 10 or more, being an important value when considering that the test allows scores from 0 to 16 correct answers per relationship type. These elements account for the non-normal distribution in most of the relationship types; only the Containment, Analogy and Deictic relationships show normal distributions (see Figure 4).

To evaluate the correlations of the RAI score with the relationship types that make up the test, the Spearman-Brown correlation coefficient was used, due to the non-parametric distribution of most of the relationship types (see Table 2). It was found that there are significant positive correlations of the overall score with all relationship types. The correlation with the Quantity, Temporal and Containment relationships stand out as presenting the highest correlations; on the other hand, the Mathematical relationships present the lowest correlation with a value of 0.41. Opposition has significant moderate to weak positive correlations with the other relationship types, except Mathematical; Temporal (0.49) stands out as the highest correlation among the relationship types. Difference also shows significant moderate to weak positive correlations with several relationship types, excluding Temporal and Deictic. Quantity has the strongest correlation with RAI (0.76), it also correlates significantly with Opposition, Difference and Containment. Temporal shows significant positive correlations, highlighting the correlation with Deictic. Containment is significantly correlated with Quantity and Temporal. Analogy shows a significant positive correlation with Difference (0.37; p <.05) and Containment (0.38; p <.05), but with a lower strength than other relationship types. Deictic has significant correlations with Quantity and Temporal. Mathematical shows lower significant correlations with Difference and Quantity, being the relationship type with the lowest correlation with other relationship types and with the overall test score

The RAI has a Cronbach's alpha of 0.88, which indicates high internal consistency, suggesting that the items consistently and reliably measure the same construct. The standard error of the *alpha* is 0.03, which is low, reflecting a high precision in this measure of consistency. Although the mean of the inter-item correlations is robust (0.67), suggesting a strong overall relationship between items, the median of these correlations is surprisingly low (0.05), which could point to the presence of items that do not correlate well with each other. Additionally, reliability was estimated by means of split-half reliability by dividing the scores obtained randomly, the result of Spearman's correlation test between the scores is -0.09 (p= .79) the correlation is very weak and negative, the high value indicates that there is no linear relationship, and p's high value indicates that there is no linear relationship between the scores of the two halves.

By calculating the Kruskal-Wallis test, a significant difference was found (H= 88.50; df= 7, p= .25 e -15), indicating statistically significant differences in the medians of the relationship types evaluated within the RAI with respect to the others. Given such

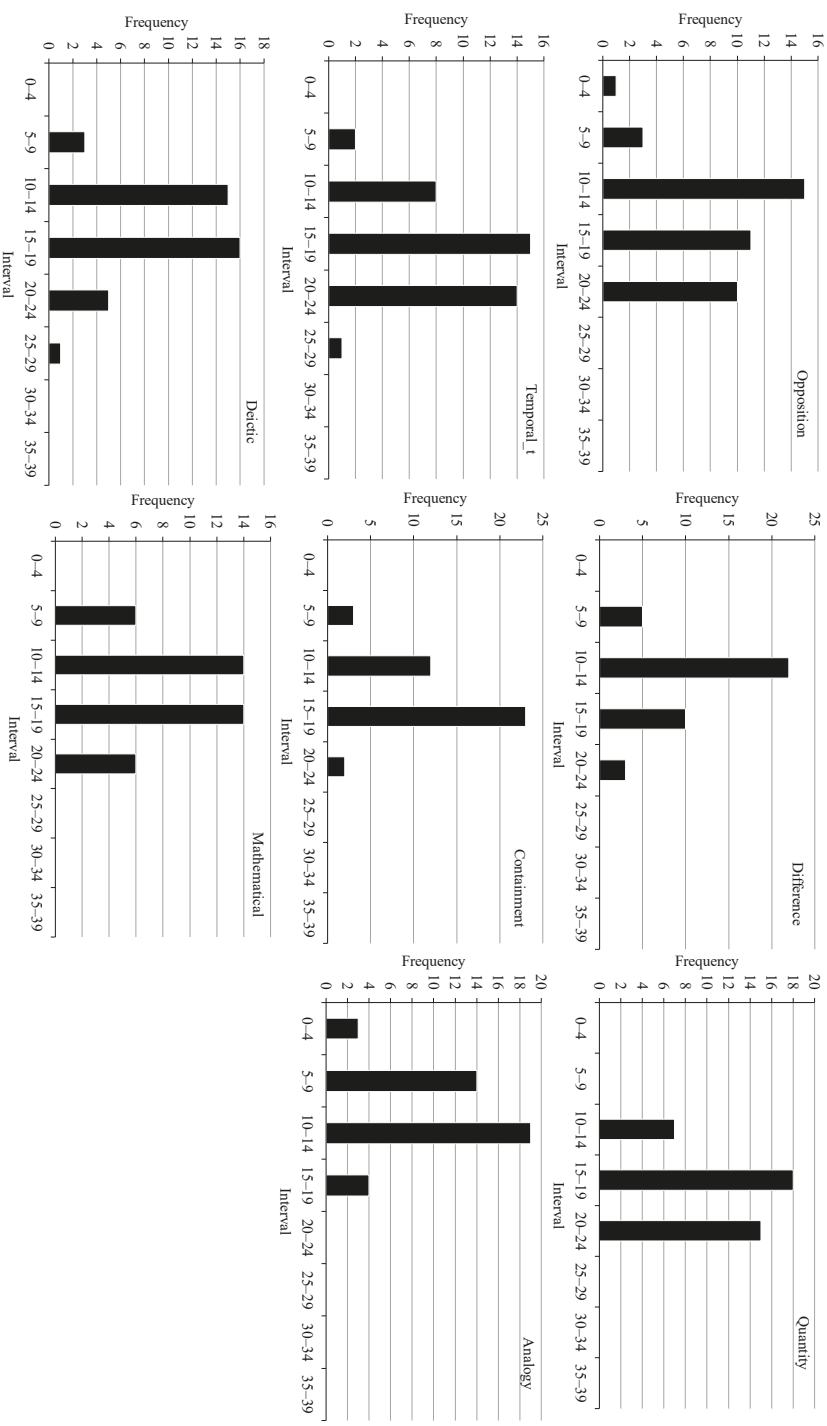

Figure 4. Histogram of the scores obtained by type of RAI relationship.The histograms displays the distribution of correct responses by type of relational task in the RAI.

Table 2. Matrix of internal correlations of the RAI test.

Relationship types	1	2	3	4	5	6	7	8
1. Opposition	-							
CI	-							
2. Difference	0.39*	-						
CI	0.09 0.63	-						
3. Quantities	0.36*	0.41**	-					
CI	0.058 0.61	0.11 0.64	-					
4. Temporal	0.49**	0.18	0.47**	-				
CI	0.28 0.70	-0.14 0.47	0.19 0.68	-				
5. Containment	0.35*	0.38*	0.48**	0.50**	-			
CI	0.04 0.60	0.08 0.62	0.20 0.69	0.22 0.70	-			
6. Analogy	0.26	0.37*	0.30	0.27	0.38*	-		
CI	-0.05 0.53	0.07 0.62	-0.01 0.56	-0.04 0.54	0.08 0.62	-		
7. Deictic	0.23	0.26	0.38*	0.51**	0.30	0.22	-	
CI	-0.09 0.50	-0.06 0.53	0.08 0.62	0.24 0.71	-0.02 0.56	-0.10 0.5	-	
8. Mathematical	0.15	0.36*	0.31*	0.11	0.21	0.19	0.02	-
CI	-0.17 0.44	0.06 0.60	0.00 0.57	-0.21 0.41	-0.11 0.49	-0.13 0.47	-0.30 0.33	-
9. RAI	0.64**	0.59**	0.76**	0.73**	0.74**	0.54**	0.55**	0.41**
CI	0.42 0.80	0.34 0.76	0.58 0.87	0.54 0.85	0.55 0.85	0.27 0.73	0.28 0.73	0.12 0.64

Notes: CI= Confident Interval 95%; *= p <.05; **= p <.01.

a low p-value, one can safely state the existence of significant differences between the relationship types. The post hoc analysis using Dunn's test, adjusted by Bonferroni's method, revealed several statistically significant individual comparisons between blocks, including Analogy with Quantitaty, Containment and Difference; Quantitaty with Deictic, Oppositional and Mathematical; Containment with Deictic, Mathematical and Opposition; as well as Difference with Deictic, Mathematical, Opposition and Temporal (see Table 3).

Table 3. Bonferroni adjusted *p* values of the comparison between relationship types

Relationship types	1	2	3	4	5	6	7
1. Opposition	-						
2. Difference	.00*	-					
3. Quantities	.00*	.89	-				
4. Temporal	1.00	.00*	.18	-			
5. Containment	.00*	1.00	1.00	.07	-		
6. Analogy	1.00	.00*	.02*	1.00	.00*	-	
7. Deictic	1.00	.00*	.01*	1.00	.00*	1.00	-
8. Mathematical	1.00	.00*	.00*	.06	.00*	.54	.70

Note: *= p <.05.

A *M* Difficulty Index of 0.67 with a *SD* of 0.16 was estimated, with a minimum value of 0.1 and a maximum of 0.92. The average values of the difficulty index were calculated by type of relationship: Difference is the type of relationship with the highest Difficulty Index (0.85), the lowest was Mathematical (0.53); maintaining in all types of relationship difficulty values from medium to high (see Table 4). An average value of the Discrimination Index of 0.23 was calculated, this value is within the range of acceptable discrimination, but close to the lower limit, the RAI in general presents the capacity to distinguish between participants with different levels of relational ability but may not be particularly powerful in this aspect (see Appendix C). As for the evaluation by type of relationship, Difference presents the best average value of the Discrimination Index (0.33), Mathematical stands out with the lowest value (0.09) (see Table 4), and

it is also the type of relationship that presents the most items with poor discrimination, presenting 7 out of 16 items with negative correlations (see Appendix C).

Table 4. Average difficulty and discrimination indices by type of RAI relationship.

Type of relationship	Difficulty index	Discrimination index
Opposition	0.56	0.25
Difference	0.85	0.33
Quantity	0.76	0.27
Temporal	0.65	0.28
Containment	0.78	0.28
Analogy	0.62	0.14
Deictic	0.62	0.18
Mathematical	0.53	0.09

The WAIS-IV scores show an increase in the *M*= 113.13, *Med*= 114.5 and a *SD*= 21.17. The *Standard Error* of the Mean was estimated at 3.34 (see Figure 5). When the normality of the distribution was evaluated by means of the Shapiro-Wilk test, the following results were obtained *W*= 0.99 *p*= .64 which suggests a normal distribution of the data, this can be visually verified in Figure 6.

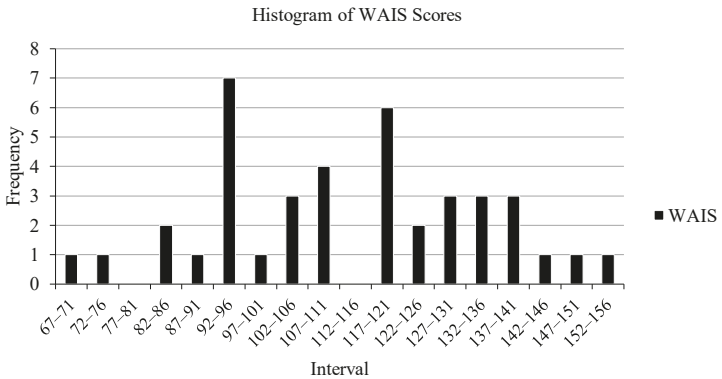

Figure 5. Histogram of the scores obtained in the WAIS-IV scale.

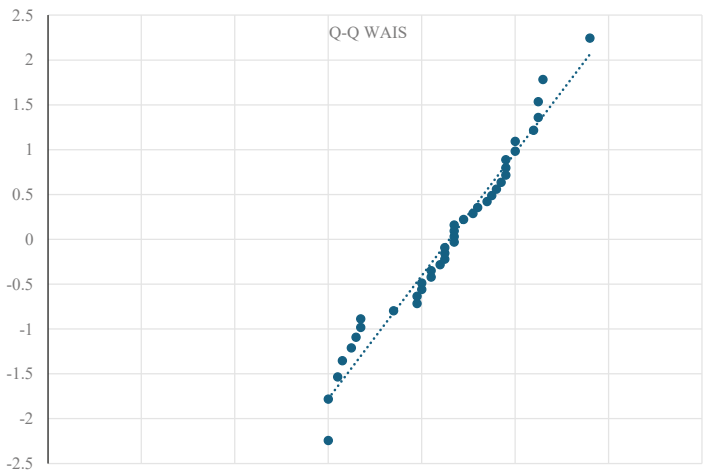

Figure 6. Q-Q plot of the scores obtained on the WAIS-IV scale.

The scores obtained in the G Factor test have a *M* of 22.97 and a *Median* of 24. The dispersion of the scores, reflected in a standard deviation of 4.23, shows a wide range of results among the participants, with scores ranging from 15 to 30 correct answers (see Figure 7). This variability is specified by a standard error of 0.67. According to the Shapiro-Wilk test, with a value of 0.95 and a value of *p* just above the significance threshold (*p*= .05) the scores could be considered as approximately normally distributed (see Figure 8).

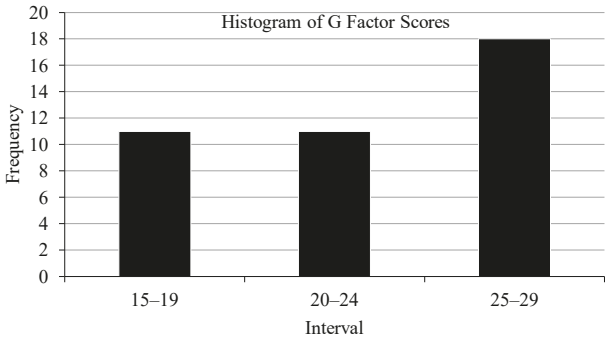

Figure 7. Histogram of the scores obtained in the G Factor test.

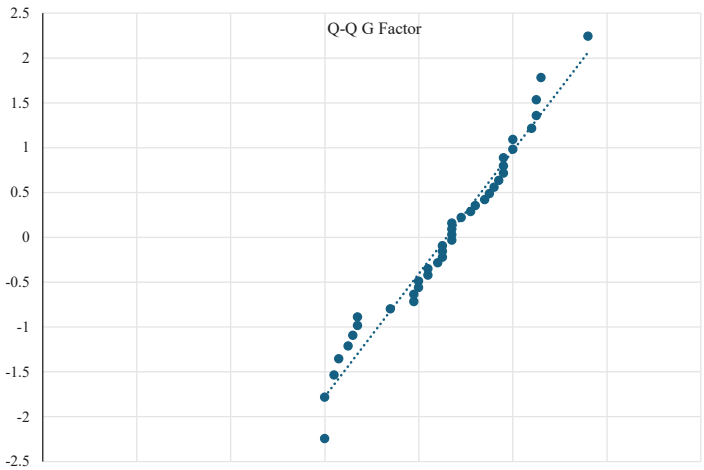

Figure 8. Q-Q plot of the scores obtained on the G Factor test.

The correlation between the RAI and the WAIS-IV scale was evaluated, with a value of r^2= 0.44 (p= .005; CI= 0.13-0.65). This indicates that approximately 44% of the variability in RAI scores can be explained by scores on the WAIS-IV scale. This suggests a moderately strong relationship between the two measures. When assessing the correlation of RAI with the subscales of the WAIS- IV test, a significant correlation was found with Block Design (r^2= 0.41 *p*= .01), suggesting that there is a moderate relationship between performance on Block Design and the RAI. With the Similarities subscale, the estimated correlation is of r^2= 0.21 *p*= .19 which is considered a low and non-significant correlation. With Matrix Reasoning there is a moderate positive

correlation (r^2= 0.38 p= .02), which shows a reliable association between the RAI and the Matrix Reasoning subtest. Finally, with Vocabulary we obtained a weak and non-significant correlation.

When calculating the correlations between the WAIS-IV scale and the relationship types, only Temporal (r^2= 0.34 p= .030) and Deictic (r^2= 0.46 p= .00) showed significant moderate correlations; the rest of the relationship types did not show significant correlations (see Table 5).

Table 5. Correlation matrix of the WAIS-IV scale and the RAI relationship types.

	WAIS	*CI*	
Opposition	0.28	-0.04	0.54
Difference	0.28	-0.03	0.55
Quantities	0.30	-0.01	0.56
Temporal	0.34*	0.03	0.59
Containment	0.18	-0.14	0.46
Analogy	0.31	-0.00	0.57
Deictic	0.46*	0.17	0.67
Mathematical	0.12	-0.2	0.41

Notes: CI= Confident Interval 95%; *= p <.05; **= p <.01.

Several correlations between different types of RAI relationships and subtests of the WAIS-IV have been found to be statistically significant (see Table 6). In particular, the Difference type shows a strong and significant correlation with Matrix Reasoning (r^2= 0.45; p= .00), indicating that the ability to make difference relations is significantly linked to visual spatial ability. Temporal also shows a positive and significant correlation with Block Design (r^2= 0.35; p= .03), suggesting that the understanding of temporal sequences may be related to visuospatial and construction skills. Analogy shows significant correlations with both Block Design (r^2= 0.36; p= .02) and Matrix Reasoning (r^2= 0.46; p= .00), implying a remarkable relationship between reasoning by analogies and both spatial and abstract skills. Deictic is particularly interesting, showing a strong correlation with Similarities (r^2= 0.49; p= .00), which could reflect a connection between spatial

Table 6. Correlation matrix of WAIS-IV subscales and RAI relationship types.

	Block Design	Similarities	Matrix Reasoning	Vocabulary
Opposition	0.27	0.16	0.27	0.21
CI	-0.05	-0.16	-0.04	-0.18
	0.54	0.45	0.54	0.49
Difference	0.20	0.27	0.45**	0.21
CI	-0.12	-0.10	0.16	-0.11
	0.48	0.50	0.67	0.49
Quantities	0.31	0.10	0.30	0.15
CI	-0.01	-0.22	-0.01	-0.17
	0.56	0.40	0.56	0.44
Temporal	0.35*	0.29	0.29	0.19
CI	0.041	-0.04	-0.03	-0.13
	0.60	0.54	0.55	0.48
Containment	0.21	0.12	0.18	0.20
CI	-0.11	-0.20	-0.14	-0.12
	0.49	0.42	0.46	0.48
Analogy	0.36*	-0.13	0.46**	0.17
CI	0.05	-0.42	0.17	-0.15
	0.60	0.19	0.67	0.46
Deictic	0.32*	0.49**	0.22	0.43**
CI	0.01	0.21	-0.10	0.13
	0.57	0.70	0.50	0.65
Mathematical	0.15	-0.02	0.14	0.04
CI	-0.17	-0.33	-0.18	-0.28
	0.44	0.29	0.43	0.35

Notes: CI= Confident Interval 95%; *= p <.05; **= p <.01.

or personal relationship understanding and verbal reasoning. In addition, Deictic has a notable correlation with Vocabulary (r^2= 0.43; p= .01), suggesting that the use and understanding of deictic terms is significantly associated with a broader vocabulary. The rest of correlations were not significant (see Table 6).

The correlation between the RAI and the G-Factor was assessed, yielding a value of r^2= 0.40 (p= .01; *CI*= 0.11-0.64). This indicates that approximately 39% of the variability in the RAI scores can be explained by the scores of the G Factor. By calculating the correlations between the G-Factor scale and the relationship types of Temporal (r^2= 0.48; p= .00), Containment (r^2= 0.40; p= .01) and Deictic (r^2= 0.386; p= .014) showed significant moderate correlations; the other correlations evaluated did not reach levels of statistical significance (see Table 7).

Table 7. Correlation matrix of the G Factor test and RAI relationship types.

	G Factor	*CI*	
Opposition	0.23	-0.04	0.54
Difference	0.20	-0.03	0.55
Quantities	0.15	-0.01	0.56
Temporal	0.48*	0.03	0.59
Containment	0.40*	-0.14	0.46
Analogy	0.27	-0.00	0.57
Deictic	0.39*	0.17	0.67
Mathematical	0.02	-0.20	0.42

Notes: CI= Confident Interval 95%; *= p <.05; **= p <.01

Discussion

This study aimed to describe the results of the Spanish adaptation of the RAI in terms of its psychometric properties; the test was administered to Mexican students of the FES Iztacala UNAM and the results were evaluated in terms of the RAI, as well as the WAIS-IV and G Factor tests as indicators of concurrent validity.

The estimated results show a normal distribution of the correct responses to the RAI, with most of the sample concentrating between 79 and 97; when evaluating the distribution of response times per item, these show a distribution with a better normal fit compared to the RAI scores; this suggests that response times can be considered as an additional parameter in the assessment of relational ability. Response times (M= 15.4 seconds) were consistent with what was reported by Ruiz *et alii* (2022) even though the version of the RAI used in the present study (Cummins *et alii*, 2023) comprise a greater number of relationship types and consequently a greater number of questions. It is worth noting that the inclusion of more items and the complexity that can add different types of relationships has no effect on average response times.

The results found by type of relationship remain in a range between 8.5 and 13.65 correct answers, being Mathematical the type of relationship with the lowest score and Differences with the highest; considering that each type of relationship was evaluated with 16 items for each one, we can affirm the existence of an important variability among the evaluated relationships, however, this variability does not respond to a progressive increase in the complexity or difficulty of the types of relationship as pointed out by Cassidy *et alii* (2016). Opposition was the first relationship type presented and the second with the lowest score, the next to be presented was difference, with the highest score the similarity in the items of these two relationship types (see Appendix B) combined

with the abrupt increase in the correct answers, leads us to consider the existence of a carry-over effect between these relationship types.

Analogy, Deictic and Mathematical can be assumed to be the most complex relationship types, since they comprise a greater number of stimuli and relationships between them per item. This is consistent with the number of correct answers, since they are the relationship types with the lowest scores (except for Opposition). Of these, Analogy and Deictic present a normal distribution, as well as scores similar to each other, with the exception of having opposite biases. Even when the RAI score presents a normal distribution, five out of eight types of relationships are not normally distributed. These data are similar to those reported by Cummins (2023) so, beyond global considerations, it is important to consider the assessment by relationship type.

Even when there is a difference in the distributions, all the relationship types present significant, positive and moderate to strong correlations with the overall RAI score, highlighting that Mathematical presents the lowest correlation. The correlations between relationship types show moderate to low significant relationships, with Quantities being the relationship type that correlates best with the rest of the relationship types and Mathematical the least correlated with other types; this is consistent with Cummins *et alii* (2023) regarding the type of relationship that presents the best correlation values, even when these authors report a different relationship as the one that presents the lowest association (Analogy), in both studies the complex relations are the ones that present the lowest levels of correlation. This may be due to the fact that the scores obtained in these types of relationships are the lowest.

Positive correlations by relationship type and a high Cronbach's alpha estimate (0.88) indicate high internal consistency. However, the Spearman correlation estimate for the split halves (-0.09 p= .69) is non-significant and negative, indicating that the RAI lacks reliability, as found by Cummins (2023). This could be contradictory with the high Cronbach's alpha, so it is important to consider additional factors such as sample size, although the effect-size analysis shows that it is an appropriate sample size for estimating correlations, splitting the sample in halves leaves us with a sample in which the correlation results may not be stable and could be affected by anomalies, by variability of responses or items that do not correlate well with each other. If some items have low variance or do not discriminate well, they could contribute negatively to split-half correlation, despite Cronbach's alpha being high, which may be driven by a set of highly correlated items, while other items are not as well aligned. This may not significantly affect Cronbach's alpha, but it does affect split-half correlation.

When comparing the means of the different types of relationship, it is observed that there are practically identical types of relationship, so that the evaluation of these types of relationship as a whole may be contributing little to the estimation of relational ability. Additionally, there are relationship types with very low difficulty indexes (eg. Difference) and the complex relationship types maintain intermediate values; in the discrimination analysis poor discrimination values are observed.

The RAI does not maintain the incremental difficulty with which it was theoretically designed (Cassidy *et alii*, 2016) either by type of relationship or considering the items that make up the assessment of each type of relationship, this is consistent with that reported by Cummins (2023). The latter suggests that one of the ways to improve the properties of the RAI in terms of discrimination is to increase the number of items per type of relationship, as suggested by Colbert *et alii* (2017); however, this may result in an excessively long test. Another proposal by Cummins (2023) is to review the levels

of difficulty by type of relationship in order to make adjustments considering only those items that present greater difficulty.

Based on the results found, it is suggested to adjust the difficulty of the items considering the complexity of the types of relationships evaluated, eliminating items and types of relationships that contribute little to the discrimination of the relational evaluation or do not present significant differences in contrast with others that present better discrimination values. This gives rise to the possibility of a more concrete test that can present a better consistency in general terms. It is also necessary to consider the order of presentation of the types of relationship, if these do not present an incremental difficulty as theoretically proposed in the design of the first version of the test; one option is the randomized presentation of the types of relationship or of the items in general, avoiding discrepancies due to the order of presentation, such as the possible carry-over effect between Opposition and Difference found in this study.

When evaluating the concurrent validity with the WAIS-IV and G Factor tests, the RAI shows moderate correlations, which allows us to give external validity to the RAI in a global manner. When considering the correlations with the types of relationships, it stands out that only the Temporal and Deictic relationships present significant correlations with the WAIS-IV test. It is important to add that only Deictic presents a considerable relationship with the verbal skills of the WAIS-IV, and Analogy is the type of relationship that more strongly associates to the non-verbal skills of the test. This contrasts with the correlations with G Factor, since Analogy shows no significant correlation with the test, even though it also measures nonverbal skills. Temporal, Deictic and Containment are the relationship types that correlate significantly with the G Factor.

The Relational Abilities Index (RAI) is a fundamental instrument for evaluating relational skills within the framework of Relational Frame Theory. Although it has strong foundational potential, substantial improvements are needed to increase its experimental applicability, especially regarding reliability and discrimination capacity. The complexity and arrangement of item types, as well as the sequence in which they are presented, require meticulous adjustments in test construction.

Although the present study utilized a sufficient sample size for initial observations. It is insufficient for conducting more intricate analyses, such as factor analysis or regression models, which could provide more comprehensive insights into the structural composition of the RAI and the relative impact of various relational types. Therefore, it is imperative to conduct additional research with larger sample sizes, not only to refine the RAI's design based on these insights but also to enhance our comprehension of how relational skills training may impact broader cognitive constructs beyond IQ.

The psychometric assessment of the RAI adaptation for the Mexican population indicates its potential utility in measuring relational intelligence, as evidenced by moderate correlations with WAIS-IV subtests focused on visuospatial and abstract reasoning. Nonetheless, the noteworthy discrepancies in reliability, exemplified by low split-half reliability in contrast to high internal consistency, necessitate supplementary investigations. These studies should aim to improve the structure of the RAI and verify its validity, thereby guaranteeing its efficacy and adaptability in diverse educational and clinical settings. This continuous evaluation and adjustment will be vital to maintaining the RAI's relevance and efficacy in assessing relational skills.

This article significantly improves our understanding of the psychometric properties of the Relational Abilities Index (RAI) within the Mexican population. Through careful adaptation and rigorous validation procedures, the study not only demonstrates the potential

utility of the RAI for evaluating relational intelligence but also bridges the gap between relational abilities and conventional intelligence measures. The findings make a significant contribution to the fields of educational and clinical psychology by demonstrating the significance of relational skills in cognitive assessments and recommending avenues for enhancing cognitive training programs. The study's methodological rigor and comprehensive analysis serve as a valuable model for future research in psychometric evaluation, promising to inspire further studies that could broaden the application of the RAI across diverse educational and cultural settings.

References

Cassidy S, Roche B, & Hayes SC (2011). A relational frame training intervention to raise intelligence quotients: A pilot study. *The Psychological Record, 61*, 173-198. Doi: 10.1007/BF03395755

Cassidy S, Roche B, Colbert D, Stewart I, & Grey IM (2016). A relational frame skills training intervention to increase general intelligence and scholastic aptitude. *Learning and Individual Differences*, *47*, 222-235. Doi: 10.1016/j.lindif.2016.03.001

Cattell R, Cattell A & Weiss R (2017). *FG-R: G Factor-R. Test de inteligencia no verbal-Revisado*. Madrid: TEA.

Colbert D, Dobutowitsch M, Roche B, & Brophy C (2017). The proxy-measurement of intelligence quotients using a relational skills abilities index. *Learning and Individual Differences*, *57*, 114-122. Doi: 10.1016/j.lindif.2017.03.010

Colbert D, Malone A, Barrett S, & Roche B (2020). The Relational Abilities Index+: Initial Validation of a Functionally Understood Proxy Measure for Intelligence. *Perspectives on Behavior Science*, *43*, 189-213. Doi: 10.1007/s40614-019-00197-z

Colbert D, Tyndall I, Roche B, & Cassidy S (2018). Can SMART Training Really Increase Intelligence? A Replication Study. *Journal of Behavioral Education*, *27*, 509-531. Doi: 10.1007/s10864-018-9302-2

Cummins J. (2023). On the measurement of relational responding. *Journal of Contextual Behavioral Science*, *30*, 155-168. Doi: 10.1016/j.jcbs.2023.10.003

Cummins J, Nevejans M, Colbert D, & De Houwer J (2023). On the structure of relational responding. *Journal of Contextual Behavioral Science*, *27*, 16-25. Doi: 10.1016/j.jcbs.2022.11.006

de Houwer J, Barnes-Holmes D, & Moors A (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin and Review*, *20*, 631-642. Doi: 10.3758/s13423-013-0386-3

Flanagan D (2014). The Cattell-Horn-Carroll theory of cognitive abilities. In CR Reynolds & KVannest J (Eds.), *Encyclopedia of Special Education* (pp. 368-373). Hoboken, NJ: John Wiley & Sons.

Fryling M, Rehfeldt R, Tarbox J, & Hayes L (2020). *Applied Behavior Analysis of Language & Cognition*. Oackland, CA: New Harbinger

Faul F, Erdfelder E, Buchner A, & Lang AG (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

González Velázquez MR, Aragón Borja LE, & Silva Rodríguez A (2000). Baremación del test de inteligencia G Factor de Cattell, en la zona metropolitana de la ciudad de México. *Psicothema*, *12*, 275-278.

Hughes S & Barnes-Holmes D (2015). Relational Frame Theory. In RD Zettle, SC Hayes, D Barnes-Holmes, & A Biglan (Eds.), *The Wiley Handbook of Contextual Behavioral Science* (pp. 179-226). Hoboken, NJ: John Wiley & Sons.

May R J, Tyndall I, McTiernan A, Roderique-Davies G, & McLoughlin S (2022). The impact of the SMART program on cognitive and academic skills: A systematic review and meta-analysis. *British Journal of Educational Technology*, *53*, 1244-1261. Doi: 10.1111/bjet.13192

Mulhern T (2022). *Relational Frame Theory: Made Simple*. Cham: Springer International Publishing.

Ruiz D, Cepeda-Islas M., Hernández-Miranda L, Hickman-Rodríguez H, & Balderas-Trejo, G (2022). Relational abilities index: An experimental study of the procedure at different trial durations. *Behavioural Processes*, *203*. Doi: 10.1016/j.beproc.2022.104766

Sánchez-Escobedo P. (2015). Challenges Measuring Intelligence in Mexico: Lessons Learned from the Standardization of the Wechsler Scales. *Educación y Ciencia, 4*, 65-79.

Schneider WJ, & Newman, DA (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resource Management Review*, *25*, 12-27. Doi: 10.1016/j.hrmr.2014.09.004

Úbeda R, Fuentes I, & Dasí C (2016). Revisión de las formas abreviadas de la Escala de Inteligencia de Wechsler para Adultos. *Psychology, Society & Education*, *8*, 81-92. Doi: 10.25115/psye.v8i1.549

Received, Juny 6, 2020
Final Acceptance, October 17, 2020

APPENDIX A

Sequence of relational skill probes and relational skill testing stages employed in Cassidy *et alii* (2016)

#	*N*	RP	PD	RQ	QD	Nodes
1	1	Same	Forwards	Mixed	Mixed	0
2	1	Opposite	Forwards	Mixed	Mixed	0
3	2	Both same	Forwards	Same	Forwards	0, 1
4	2	Both same	Forwards	Same	Backwards	0, 1
5	2	Both same	Forwards	Same	Mixed	1
6	2	Both	Forwards	Same	Forwards	0, 1
7	2	Opposite Both	Forwards	Same	Backwards	0, 1
8	2	Opposite Both	Forwards	Same	Mixed	1
9	2	Opposite BS	Forwards	Mixed	Mixed	1
10	2	Both Same	Mixed	Same	Mixed	1
11	2	Both Same	Mixed	Opposite	Mixed	1
12	2	Both Same	Mixed	Mixed	Mixed	1
13	2	Both	Forwards	Mixed	Mixed	1
14	2	Opposite Both	Forwards	Mixed	Mixed	0, 1
15	2	Opposite Both	Mixed	Same	Mixed	0, 1
		Opposite	-	-	-	-
16	2	Both	Mixed	Opposite	Mixed	0, 1
17	2	Opposite Both	Mixed	Mixed	Mixed	0, 1
18	2	Opposite Mixed	Forwards	Same	Mixed	1
19	2	Mixed	Forwards	Opposite	Mixed	1
20	2	Mixed	Forwards	Mixed	Mixed	0, 1
21	2	Mixed	Mixed	Mixed	Mixed	0, 1
22	3	Same	Forwards	Mixed	Mixed	1
23	3	Same	Forwards	Mixed	Mixed	1, 2
24	3	Opposite	Forwards	Mixed	Mixed	1
25	3	Opposite	Forwards	Mixed	Mixed	1, 2
26	3	Mixed 1	Forwards	Mixed	Mixed	1
27	3	(S/S/O) Mix2	Forwards	Mixed	Mixed	1, 2
28	3	(S/S/O) Mix3	Forwards	Mixed	Mixed	1, 2
29	3	(S/S/O) Mix4	Forwards	Mixed	Mixed	1, 2
30	1	(O/O/S) More	Forwards	Mixed	Mixed	0
31	1	Less	Forwards	Mixed	Mixed	0
32	2	Both more	Forwards	More	Forwards	0, 1
33	2	Both more	Forwards	More	Mixed	0, 1
34	2	Both more	Forwards	Less	Mixed	0, 1
35	2	Both more	Forwards	Mixed	Mixed	1
36	2	Both less	Forwards	Less	Forwards	1
37	2	Both less	Forwards	Less	Mixed	0, 1
38	2	Both less	Forwards	More	Mixed	0, 1
39	2	Both less	Forwards	Mixed	Mixed	1
40	2	Both more	Mixed	More	Mixed	0, 1
41	2	Both more	Mixed	Less	Mixed	0, 1
42	2	Both more	Mixed	Mixed	Mixed	1
43	2	Both less	Mixed	Less	Mixed	0, 1
44	2	Both less	Mixed	More	Mixed	0, 1
45	2	Both less	Mixed	Mixed	Mixed	1
46	3	All more	Forwards	Mixed	Mixed	0, 1
47	3	All more	Forwards	Mixed	Mixed	1, 2
48	3	All less	Forwards	Mixed	Mixed	0, 1
49	3	All less	Forwards	Mixed	Mixed	1, 2
50	3	All more	Mixed	Mixed	Mixed	1
51	3	All more	Mixed	Mixed	Mixed	2
52	3	All more	Mixed	Mixed	Mixed	0, 1, 2
53	3	All less	Mixed	Mixed	Mixed	1
54	3	All less	Mixed	Mixed	Mixed	2

Notes: Mixed 1= Mixed type 1; Nodes= Number of nodes; Opposite BS= Opposite Both Same; PD= Premises Direction; QD= Question Direction; RQ= Relations in Questions; RP= Relations in premises; (S/S/O) Mix2 = (S/S/O) Mixed type 2; (S/S/O) Mix3= (S/O/O) Mixed type 3; (S/S/O) Mix4= (O/S/S) Mixed type 4.

APPENDIX B

Examples of RAI reagents applied in Cummins *et alii* (2023)

APPENDIX C

Difficulty and discrimination indices per RAI item. (Type= Type of relationship)

Type	Item	Difficulty	Discrimination
Opposition	RAI1	0.625	0.455
	RAI2	0.500	0.371
	RAI3	0.675	0.395
	RAI4	0.775	0.584
	RAI5	0.625	0.365
	RAI6	0.525	0.383
	RAI7	0.625	0.384
	RAI8	0.725	0.029
	RAI9	0.700	0.462
	RAI10	0.475	0.372
	RAI11	0.375	0.168
	RAI12	0.350	0.059
	RAI13	0.575	0.075
	RAI14	0.500	0.116
	RAI15	0.400	-0.334
	RAI16	0.450	0.134
Difference	RAI17	0.875	0.374
	RAI18	0.900	0.190
	RAI19	0.875	0.136
	RAI20	0.850	0.187
	RAI21	0.875	0.076
	RAI22	0.825	0.408
	RAI23	0.800	0.133
	RAI24	0.825	0.328
	RAI25	0.825	0.523
	RAI26	0.825	0.375
	RAI27	0.800	0.503
	RAI28	0.875	0.412
	RAI29	0.925	0.189
	RAI30	0.850	0.287
	RAI31	0.825	0.605
	RAI32	0.900	0.538
Quantities	RAI33	0.775	0.566
	RAI34	0.750	0.230
	RAI35	0.850	0.479
	RAI36	0.875	0.163
	RAI37	0.750	0.234
	RAI38	0.800	0.040
	RAI39	0.650	0.421
	RAI40	0.750	0.616
	RAI41	0.750	0.209
	RAI42	0.825	0.290
	RAI43	0.725	0.001
	RAI44	0.725	0.321
	RAI45	0.700	0.582
	RAI46	0.800	-0.027
	RAI47	0.750	0.093
	RAI48	0.725	0.124
Temporal	RAI49	0.550	0.178
	RAI50	0.650	0.375
	RAI51	0.575	0.064
	RAI52	0.600	0.230
	RAI53	0.775	0.409
	RAI54	0.600	0.367
	RAI55	0.675	0.349
	RAI56	0.675	0.384
	RAI57	0.625	0.116
	RAI58	0.625	0.175
	RAI59	0.725	0.200
	RAI60	0.525	0.231
	RAI61	0.725	0.558
	RAI62	0.575	0.256
	RAI63	0.825	0.332
	RAI64	0.725	0.261
Containment	RAI65	0.725	0.204
	RAI66	0.875	0.190
	RAI67	0.800	0.151
	RAI68	0.800	0.241
	RAI69	0.775	0.379
	RAI70	0.825	0.191
	RAI71	0.875	0.157
	RAI72	0.725	0.509
	RAI73	0.725	0.431
	RAI74	0.775	0.331
	RAI75	0.700	0.299
	RAI76	0.625	0.271
	RAI77	0.775	0.266
	RAI78	0.850	0.332
	RAI79	0.775	0.431
	RAI80	0.825	0.050
Analogy	RAI81	0.800	0.067
	RAI82	0.650	0.277
	RAI83	0.525	0.453
	RAI84	0.875	0.396
	RAI85	0.800	-0.013
	RAI86	0.675	0.450
	RAI87	0.725	0.447
	RAI88	0.725	0.257
	RAI89	0.200	0.005
	RAI90	0.850	0.007
	RAI91	0.425	-0.059
	RAI92	0.925	0.128
	RAI93	0.725	-0.102
	RAI94	0.300	0.283
	RAI95	0.100	-0.225
	RAI96	0.550	-0.111
Deictic	RAI97	0.875	0.114
	RAI98	0.850	0.137
	RAI99	0.775	0.219
	RAI100	0.500	0.238
	RAI101	0.775	0.323
	RAI102	0.700	0.158
	RAI103	0.500	0.173
	RAI104	0.550	0.433
	RAI105	0.600	0.066
	RAI106	0.600	-0.014
	RAI107	0.525	0.123
	RAI108	0.600	0.245
	RAI109	0.575	0.097
	RAI110	0.600	0.238
	RAI111	0.425	0.063
	RAI112	0.475	0.194
Mathematical	RAI113	0.575	0.195
	RAI114	0.400	-0.152
	RAI115	0.400	0.173
	RAI116	0.400	0.480
	RAI117	0.425	0.157
	RAI118	0.600	-0.222
	RAI119	0.425	-0.268
	RAI120	0.500	-0.157
	RAI121	0.750	0.188
	RAI122	0.625	0.271
	RAI123	0.400	0.276
	RAI124	0.525	-0.019
	RAI125	0.650	0.387
	RAI126	0.675	-0.003
	RAI127	0.625	0.212
	RAI128	0.525	-0.100