

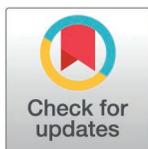
RESEARCH ARTICLE

Normative Data for Learning and Memory Test (TAMV-I) in Latin American and Spanish Children: An item response theory and linear mixed models approach

Eliana María Fuentes Mendoza^{1,2}, Laiene Olabarrieta-Landa^{1,3}, Alberto Rodríguez-Lorenzana¹, Guido Mascialino⁴, Esperanza Vergara-Moragues⁵, Carlos José de los Reyes-Aragón⁶, Natalia Albaladejo-Blázquez⁷, Natalia Cadavid-Ruiz⁸, María José Irias Escher^{9,10}, Juan Carlos Arango-Lasprilla^{11,12}, Erick Orozco-Acosta^{13,14*}, Diego Rivera^{1,3*}

1 Department of Health Science, Public University of Navarre, Pamplona, Spain, **2** Centro de Experimentación e Investigación Aplicada en Psicometría y Evaluación, CEI-TEST, Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras, **3** Instituto de Investigación Sanitaria de Navarra (IdiSNA), Pamplona, Spain, **4** Escuela de Psicología y Educación, Universidad de Las Américas, Quito, Ecuador, **5** Departamento de Psicología, Universidad de Cádiz, Cádiz, Spain, **6** Department of psychology, Universidad del Norte, Barranquilla, Colombia, **7** Department of Health Psychology, University of Alicante, Alicante, Spain, **8** Pontificia Universidad Javeriana, Cali, Colombia, **9** Escuela de Ciencias Psicológicas, Facultad de Ciencias Sociales, Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras, **10** Grupo de Investigación en Neurociencias Aplicadas, Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras, **11** Department of Cell Biology and Histology, University of the Basque Country, UPV/EHU, Leioa, Spain, **12** IKERBASQUE, Basque Foundation for Science, Bilbao, Spain, **13** Engineering Faculty, Universidad Simón Bolívar, Barranquilla, Colombia, **14** Life Sciences Research Center, Universidad Simón Bolívar, Barranquilla, Colombia

* erick.orozco@unisimon.edu.co (EO-A); diego.rivera@unavarra.es (DR)



OPEN ACCESS

Citation: Fuentes Mendoza EM, Olabarrieta-Landa L, Rodríguez-Lorenzana A, Mascialino G, Vergara-Moragues E, de los Reyes-Aragón CJ, et al. (2026) Normative Data for Learning and Memory Test (TAMV-I) in Latin American and Spanish Children: An item response theory and linear mixed models approach. PLoS One 21(2): e0341237. <https://doi.org/10.1371/journal.pone.0341237>

Editor: Alejandro Botero Carvajal, Universidad Santiago de Cali, COLOMBIA

Received: January 23, 2025

Accepted: January 5, 2026

Published: February 18, 2026

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0341237>

Copyright: © 2026 Fuentes Mendoza et al. This is an open access article distributed under the

Abstract

Robust normative data for pediatric learning and memory tests in Spanish-speaking populations are scarce, and existing approaches often rely on univariate methods that overlook item-level properties and inter-trial dependencies. The aim was to evaluate the item parameters of the TAMV-I using Item Response Theory (IRT) and to generate covariate-adjusted normative data through Linear Mixed Models (LMM). We hypothesized that the 2-parameter logistic (2PL) model would outperform the Rasch model and that demographic and contextual factors would show significant interactions influencing test performance. The sample consists of 1640 participants from Spain, Honduras, Ecuador, and Colombia. The inclusion criteria were being 6–17 years old, IQ ≥ 80 on TONI-2, and score < 19 on the Children’s Depression Inventory (CDI). Children with a history of neurological and/or psychiatric disorders were excluded. Item parameters were determined using the 1,2-PL model. LMM were used to evaluate the effect of sociodemographic variables (sex, age, age², mean parent years of education-MPE, country, and interactions). Norms were generated based on participant ability. As a result, the item parameters were calculated and the

terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are available within the paper and its [Supporting information](#) files.

Funding: This study was funded by the Carolina Foundation for support of the 2024 postdoctoral internship for Carlos José de los Reyes-Aragón. Other authors did not receive any specific funding for this work. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

LMM showed significant interactions for *MPE * Country*, *Age * Trial*, *Sex * Country* and *Trial * Country*. By integrating IRT with LMM, this study provides cross-national, covariate-adjusted norms for the TAMV-I, enhancing precision and clinical validity compared to previous approaches.

Introduction

Children's development relies on several factors, some of them are inherent to the individual, and others, to their environment [1–3]. Children's development is understood as a multidimensional process that includes physical, emotional, social, and cognitive dimensions, allowing the child to adaptively face environmental demands [4,5]. One of these cognitive abilities is learning and memory. Many studies have reported learning and memory difficulties in several neurodevelopmental disorders, such as attention deficit hyperactivity disorder [6,7], specific reading learning disabilities [8], written expression disorder [9], dyscalculia [10], intellectual disability [11,12], and autism spectrum disorders [13,14], among others. These learning and memory difficulties are associated with many functional impairments. Several studies have suggested that children with memory impairments may not only exhibit poor academic performance [15–18], but also social difficulties [19,20], and even some gaps in other cognitive skills, such as language [21].

Given the importance of memory and learning in childhood, many neuropsychological instruments have been developed for children. Some of these instruments were designed for the assessment of an isolated type of memory, such as working memory [22], prospective memory [23], and especially short- and long-term memory [24], in both auditory-verbal [25–27] and visual modalities [28,29]. Other instruments assess memory skills within larger and rigid protocols that evaluate additional cognitive skills [30–32].

Among the most widely used instruments in Latin America for assessing short- and long-term memory are the Rey Auditory Verbal Learning Test, the California Verbal Learning Test, the NEUROPSI, and the Spain-Complutense Verbal Learning Test, while for visual memory, the Rey-Osterrieth Complex Figure Test [24]. However, using these memory assessment instruments in children has some drawbacks. First, most instruments been designed for adult populations from the United States of America and Europe [33]. Although there have been initiatives to develop normative data for other countries [28,34–36], culturally adapted instruments are still limited [33,37]. This inaccuracy in estimating skills could be related, for example, to the type of words included in a word list and their frequency of use at different ages, in various cultures, and even, in different eras [38–40].

Furthermore, it has recently been shown that factors beyond strictly linguistic ones could affect children's cognitive performance, such as parents' educational level [41] and socioeconomic status [32], could affect children's cognitive performance, with the influence of such factors also varying between countries. In fact, according to a study conducted by Arango-Lasprilla et al. [24], which included 808 neuropsychologists

from 17 Latin American countries, more than 60% of the professionals considered the lack of normative data for their country of origin as one of the main problems of assessment instruments. Another disadvantage reported by Arango-Lasprilla et al. [24] study was the high cost of assessment instruments. Approximately 50% of the participants considered the cost of the instruments as a disadvantage of most existing tests. This finding is understandable considering that the average reported income in the study was around USD\$1500, and in some cases, a single assessment instrument can cost up to USD\$1000.

Currently, one of the main memory assessment instruments for children in Latin America is the Evaluación Neuropsicológica Infantil [Neuropsychological Evaluation for Children] (ENI). This test, developed in Mexico, was designed for Spanish-speaking populations and is, therefore, widely used in Latin America [24]. However, it can be costly for professionals, and its application is time-consuming as it does not only assess memory, but also other cognitive skills and academic abilities [42]. Additionally, the normative data were originally developed for Mexican children, so its use in other countries can lead to biased results. Only one study has developed ENI normative data for the Colombian population [31], but this study was conducted with a sample from a single region of the country. The researchers stratified the sample of 252 children by age, resulting in groups of just over 60 children. Finally, the normative data were generated through multiple analyses of variance (MANOVAs), so the estimates may be less precise than those from more current models [43].

Rivera et al. [44] published a study presenting normative data for the new Test of Verbal Learning and Memory (TAMV-I) for 9 Latin American countries (Chile, Cuba, Ecuador, Guatemala, Honduras, Mexico, Paraguay, Peru, and Puerto Rico) and Spain. This test was developed to assess the learning and memory of Spanish-speaking populations aged from 6 to 17 years, and it has shown good psychometric properties. As an open-license test with normative data for Latin America, it represents a good alternative for clinical practice. However, as with most verbal learning tests, the normative data for total learning scores, delayed recall, and recognition are typically estimated under the assumption that these scores are independent and identically distributed. According to Van der Elst et al. [45], this approach is inadequate when the scores are related, as in the case of the TAMV-I for three reasons. First, the correlated nature of the data is unsuitable for univariate analyses. Second, considering that the TAMV-I yields six scores, statistical models for several of these results need to be tested, potentially increasing type I error, reducing analysis power, and consequently biasing normative data. Finally, calculating six univariate regression models contradicts the principle of parsimony.

From a psychometric perspective, Classical Test Theory (CTT) is widely used, but it assumes constant measurement error across individuals and does not incorporate item-level parameters. Consequently, it provides only limited information about where a test measures most precisely and often relies on separate univariate adjustments for each score, which can inflate error rates and obscure inter-trial dependencies. Moreover, parameter estimates such as difficulty and discrimination in CTT are sample-dependent, which may introduce bias into the results [46]. In contrast, IRT models account for item difficulty and discrimination, allowing measurement precision to vary along the ability continuum. This framework provides richer psychometric information, supports the development of cross-national and covariate-adjusted norms, and facilitates adaptive testing designs [47]. Considering that TAMV-I produces correlated responses across successive trials and delays, combining IRT with linear mixed models (LMM) better captures within-person correlation and between-person covariates than univariate adjustments alone [48]. This methodological approach is not only statistically robust but also clinically meaningful, as it reflects the inter-trial dependencies that clinicians use to interpret performance patterns. In practice, the derived norms reduce the risk of over- or underestimating impairment, thereby improving diagnostic accuracy and guiding more targeted interventions.

To our knowledge, few studies have combined IRT with LMMs to produce cross-national, Spanish-language pediatric norms for list-learning tests. Therefore, this study aims to develop normative data for the TAMV-I test by combining item response theory (IRT) models with mixed-effects models, leveraging the strengths of both approaches to provide robust and precise normative estimates.

Methods

Participants

The original sample consisted of 1,748 children and adolescents from Spain ($n=399$), Honduras ($n=288$), Colombia ($n=457$), and Ecuador ($n=604$). Most of the sample were female (52.56%) with an average age of 11.19 (SD=3.36) and the mean parental education (MPE) was 13.32 years (SD=3.87). The final sample used for the analyses comprised 1,640 participants with complete data. The sample size for each country was subject to availability at the collaborating institutions rather than predetermined. Nevertheless, local research teams ensured balanced distributions across sex and age groups, and MPE was monitored in each subsample (see [Table 1](#)). Following Innocenti et al. [49], and assuming a 95% ($Z_{1-\alpha/2}$) confidence level and with $Z_0 = -0.954$, the expected standard error (τ) was 0.2679 for Spain, 0.3153 for Honduras, 0.2503 for Colombia, 0.2178 for Ecuador, and 0.1280 for the total sample. These values fall within adequate ranges, confirming that the achieved precision is sufficient both for the study aims and for the generation of robust and clinically meaningful normative data. Further details of the sample are available in [Table 1](#).

Table 1. Sample distribution by country, age, MPE, and sex.

Age range (Years)	Age (Years)		MPE	Sex	
	n	M (SD)	M (SD)	Girl n (%)	Boy n (%)
Colombia					
6–8	121	7.0 (0.8)	12.3 (3.6)	63 (52.1%)	58 (47.9%)
9–11	117	10.1 (0.8)	12.3 (3.7)	63 (53.8%)	54 (46.2%)
12–14	109	13.0 (0.8)	12.3 (3.9)	59 (54.1%)	50 (45.9%)
15–17	87	15.9 (0.8)	12.4 (4.2)	50 (57.5%)	37 (42.5%)
Ecuador					
6–8	142	7.0 (0.8)	13.5 (3.5)	70 (49.3%)	72 (50.7%)
9–11	144	10.1 (0.8)	13.7 (4.0)	76 (52.8%)	68 (47.2%)
12–14	136	12.9 (0.8)	13.5 (3.6)	69 (50.7%)	67 (49.3%)
15–17	135	16.0 (0.9)	12.8 (3.7)	69 (51.1%)	66 (48.9%)
Honduras					
6–8	68	7.0 (0.8)	13.2 (3.7)	36 (52.9%)	32 (47.1%)
9–11	92	10.0 (0.8)	12.7 (3.6)	46 (50.0%)	46 (50.0%)
12–14	68	13.1 (0.9)	12.0 (4.0)	39 (57.4%)	29 (42.6%)
15–17	56	15.9 (0.9)	13.7 (3.3)	33 (58.9%)	23 (41.1%)
Spain					
6–8	103	7.0 (0.8)	15.0 (3.8)	53 (51.5%)	50 (48.5%)
9–11	114	10.1 (0.8)	16.3 (3.6)	56 (49.1%)	58 (50.9%)
12–14	77	12.9 (0.8)	13.7 (4.0)	39 (50.6%)	38 (49.4%)
15–17	71	16.1 (0.8)	13.5 (3.5)	41 (57.7%)	30 (42.3%)
Total					
6–8	434	7.0 (0.8)	13.5 (3.8)	222 (51.2%)	212 (48.8%)
9–11	467	10.1 (0.8)	13.8 (4.0)	241 (51.6%)	226 (48.4%)
12–14	390	13.0 (0.8)	12.9 (3.9)	206 (52.8%)	184 (47.2%)
15–17	349	16.0 (0.8)	13.0 (3.77)	193 (55.3%)	156 (44.7%)

Note: MPE= Mean years Parents Education.

<https://doi.org/10.1371/journal.pone.0341237.t001>

To be included in this study, participants needed to meet the following inclusion criteria: a) be between 6–17 years old, b) be born in any of the four participant countries, c) an IQ ≥ 80 on the Test of Non-verbal Intelligence TONI-2 [50], and d) a score of < 19 on the Children's Depression Inventory [51]. Participants were ineligible if they reported: a) History of central nervous system disorders with neuropsychological impact (e.g., epilepsy, brain injury, multiple sclerosis), b) alcohol abuse or psychotropic substance use, c) uncontrolled systemic diseases causing cognitive issues (e.g., diabetes, hypothyroidism), d) psychiatric disorders (e.g., depression, bipolar disorder), e) severe sensory deficits affecting test performance, f) intellectual disabilities or neurodevelopmental disorders, g) pre-, peri-, or post-natal complications (e.g., hypoxia, seizures), h) having a score of > 5 on the Alcohol Use Disorders Identification Test -AUDIT-C [52] for participants 12 years of age and older, and j) using psychoactive substances such as heroin, barbiturates, amphetamines, methamphetamines, or cocaine in the last 6 months for participants 12 years of age and older.

Instrument

Verbal Learning and Memory Test (TAMV-I). The TAMV-I is a neuropsychological test that evaluates the verbal learning and memory in children, and it consists of three components: free recall, delayed recall, and recognition. Free recall involves four trials where the evaluator reads a list of 12 words (categorized under clothing, furniture, and body parts) after which the examinee is asked to recall as many words as possible. Delayed recall occurs 30 minutes after the fourth trial, where the examinee is prompted to recall all the words s/he can remember from the previous trials. In the Recognition phase, the individual is presented with a list of 48 words, including the original 12 words, along with 12 semantically related words, 12 phonologically related words, and 12 semantically unrelated words. Scoring entails awarding one point for each correctly recalled/recognized word from the original list of 12 words, resulting in a maximum score of 48 for free recall, 12 for delayed recall, and 12 for recognition [53]. In line with the test manual, all administrations were carried out using paper-and-pencil format.

Procedure

This study is part of a broader research project aimed at generating statistical normative data for various neuropsychological measures across Latin American countries and Spain. Ethical approval was obtained from the following institutions: the Education Committee at the International University of La Rioja (Spain); the Ethics Committee for Research in the Health Sciences Division of the Universidad del Norte (Colombia); the Ethics Committee for Research of the Universidad Pedagógica y Tecnológica de Colombia; the Ethics Committee for Human Research of the Universidad San Francisco de Quito (Ecuador); and the Ethics Committee for Research of the Master's Program in Infectious and Zoonotic Diseases (CEI-MIEZ, Honduras).

Data collection occurred from 03/01/2016–09/06/2017. Local research teams first established agreements with schools and high-schools in each country. Once authorization was obtained from the institutions, the project was presented to students and their families, who were invited to participate on a voluntary basis. Written informed consent was secured from all parents/guardians and participants aged 12 and older, while written assent was obtained from children under 12. The consent process detailed the study's objectives, participant rights, assessment duration and location, and contact information for the local researcher. Parent questionnaires were reviewed before the assessments, which were conducted individually in schools or universities. The neuropsychological battery lasted approximately 120 minutes and was administered in accordance with the guidelines of each test's manual. Participation was voluntary, with no financial incentives offered. Further details are available in Rivera and Arango-Lasprilla [41].

Statistical analysis

Item parameters and ability scores. To determine the item parameters (difficulty and discrimination), IRT was used. Since the data exhibited a dichotomous nature, both the Rasch model and the Two-Parameter Logistic (2PL) model were employed. Subsequently, the likelihood ratio test was used to compare nested models, and the Bayesian Information

Criterion (BIC) to determine the model that demonstrated the optimal fit that solves the problem of overfitting caused by the number of parameters in the model. The Rasch model operates under the assumption that items fluctuate exclusively according to the difficulty parameter [54]. In accordance with Rizopoulos' [54] notation, the mathematical representation for delineating the Rasch model is as follows:

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_i + \beta_z$$

where π_i is the conditional probability of providing a correct response to the i th item given \mathbf{z} , β_i represents the parameter denoting the ease of the i th item, β stands for the discrimination parameter (uniform across all items), and \mathbf{z} is the latent ability. The 2PL model estimates both difficulty and discrimination parameters for each individual item.

Once the best-fitting model was selected, item parameters were used to estimate each participant's ability score (θ). This score reflects the underlying performance level by weighting responses according to item difficulty and discrimination, providing a more accurate measure than raw totals. The procedure was applied separately for Trials 1–4, as well as for Delayed Recall and Recognition, yielding comparable and standardized estimates of ability across all test components.

Demographic effects. To examine the influence of demographic factors on ability scores (θ), we applied Linear Mixed Models (LMMs), which are well suited for repeated measures data such as the six trials of the TAMV-I (Trials 1–4, Delayed Recall, and Recognition). LMMs allow the inclusion of both fixed effects (trial, age, age², sex, MPE, country, and their second levels interactions) and random effects to account for within-subject variability. Importantly, in contrast to univariate regression, LMMs model multiple correlated outcomes jointly, meaning that all trial scores are considered within the same framework [55,56]. This can be defined using the following mathematical expression:

$$\theta_{ij} = \beta_0 + \mathbf{b}_{0i} + \beta_1 \mathbf{X}_{1ij} + \dots + \beta_k \mathbf{X}_{kij} + \epsilon_{ij},$$

where θ_{ij} represents the ability score for individual i in trial j , β_0 and β_1 are the fixed effects, \mathbf{b}_0 denotes the vector of random effects, and ϵ_{ij} represents the errors for subject i .

The Restricted Maximum Likelihood (REML) criterion serves as a measure to assess the fit of the LMM. It operates by considering the likelihood of data transformed into contrasts and offers the advantage of unbiased estimation of σ [57]. It has the advantage of estimating covariance parameters while appropriately accounting for the loss of degrees of freedom in estimation [58].

To select the optimal model containing the best predictor variables for ability scores (θ), sequential replacement selection approach was used, which iteratively replaces predictors to improve model fit. This method is computationally efficient and scalable. These models are then compared, and the optimal one is selected according to the BIC [59].

Normative data procedure. To generate normative conversions of the ability scores (θ_s) into percentile values adjusted for demographic factors we used predictions from the final LMM. First, the expected ability scores ($\hat{\theta}_i$) were computed using the final linear mixed-effects regression model: $\hat{\theta}_{ij} = \hat{\beta}_0 + \mathbf{b}_{0i} + \hat{\beta}_1 \mathbf{X}_{1ij} + \dots + \hat{\beta}_k \mathbf{X}_{kij} + \epsilon_{ij}$. Second, the cumulative probability of the observed ability estimates (θ_s) for participant i was obtained from the standard normal cumulative distribution function. Finally, this probability was multiplied by 100 to obtain the corresponding percentile rank (PR_{ij}). To facilitate understanding, the Fig 1 provides a schematic diagram of the statistical procedure used in this study.

All analyses were performed using R Project for Statistical Computing for Windows [60] with the *lme4* [61], *lmerTest* [56], and *ltm* packages [54]. The full analysis scripts are available at: <https://github.com/diegoriveraps/tamvi-irt-lmm-scripts>

Results

Item parameters and ability scores

IRT analyses compared Rasch (1PL) and 2PL models for each trial. Likelihood-ratio tests indicated that the 2PL model provided a significantly better fit than the Rasch model ($p < .001$). Consistently, lower BIC values further supported the

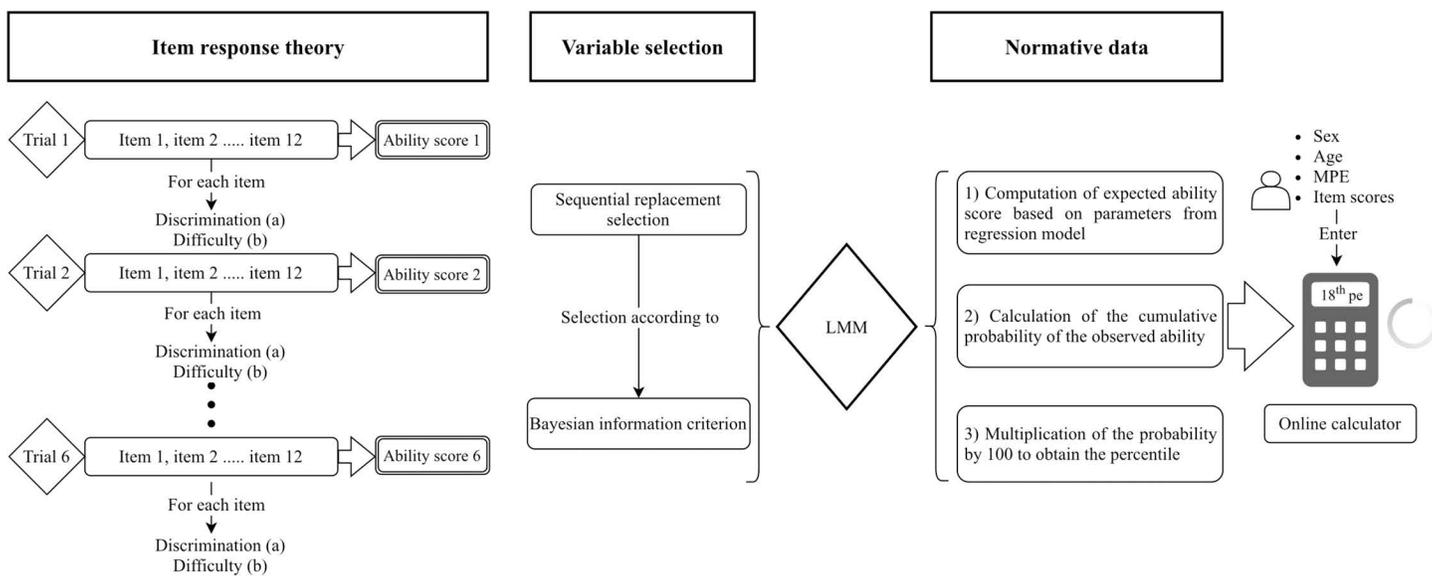


Fig 1. Workflow of item-level modeling and normative data generation for the TAMV-I.

<https://doi.org/10.1371/journal.pone.0341237.g001>

selection of the 2PL model, confirming that the additional parameter for item discrimination meaningfully improved model performance. Table 2 summarizes the estimated item parameters (discrimination and difficulty) across the six trials, allowing for the identification of those items that were most effective at differentiating between individuals with varying ability levels, as well as those that were comparatively less informative (for complete results of each trial, see S1 Appendix).

As representative examples, in Trial 4, the easiest item was Nariz [Nose] ($b = -3.17$), whereas Sillón [Armchair] was the most difficult ($b = -0.51$). Discrimination peaked for Zapato [Shoe] ($a = 1.09$) and Blusa [Blouse] ($a = 1.00$), while Sillón again showed the lowest value ($a = 0.30$). In Delayed Recall, Comedor [Dining room] emerged as the easiest ($b = -0.51$)

Table 2. Most representatively item parameter.

Trial	BIC		Discrimination		Difficulty	
	Rasch	2PL	Upper item	Lower item	Upper item	Lower item
Free recall-Trial 1	25286.57	25177.49*	Zapato [Shoe] ($a = 0.60$)	Nariz [Nose] ($a = -1.19$)	Sillón [Armchair] ($b = 72.61$)	Bufanda [Scarf] ($b = -13.68$)
Free recall-Trial 2	25643.72	25560.28*	Armario [Wardrobe] ($a = 1.75$)	Escritorio [Desk] ($a < 0.01$)	Sillón [Armchair] ($b = 2.20$)	Escritorio [Desk] ($b = -267.27$)
Free recall-Trial 3	23521.28	23496.06*	Zapato [Shoe] ($a = 1.39$)	Sillón [Armchair] ($a = 0.33$)	Sillón [Armchair] ($b = 0.35$)	Escritorio [Desk] ($b = -2.35$)
Free recall-Trial 4	21005.03	21032.14*	Zapato [Shoe] ($a = 1.09$)	Sillón [Armchair] ($a = 0.29$)	Sillón [Armchair] ($b = -0.50$)	Nariz [Nose] ($b = -3.17$)
Delayed recall-Trial 5	22737.40	22761.46*	Blusa [Blouse] ($a = 1.24$)	Nariz [Nose] ($a = 0.57$)	Sillón [Armchair] ($b = 0.60$)	Armario [Ward- robe] ($b = -2.04$)
Recognition-Trial 6	7782.25	7824.35*	Ojo [Eye] ($a = 3.20$)	Armario [Wardrobe] ($a = 1.63$)	Sillón [Armchair] ($b = -1.60$)	Nariz [Nose] ($b = -2.45$)

Note: BIC=Bayesian Information Criterion; * p -value < 0.001 for Likelihood Ratio Test. This table shows which items in each trial were the easiest, the hardest, and the most effective at distinguishing between children based on their abilities.

<https://doi.org/10.1371/journal.pone.0341237.t002>

and Sillón as the hardest ($b=0.60$), with discrimination highest for Blusa ($a=1.24$) and lowest for Nariz [Nose] ($a=0.57$). Finally, in the Recognition trial, Nariz ($b=-2.46$) and Comedor ($b=-2.21$) were among the easiest, whereas Sillón remained one of the hardest ($b=-1.60$). Discrimination values reached their maximum in this condition, with Ojo [Eye] ($a=3.20$) showing the strongest slope, closely followed by Oreja [Ear], Boca [Mouth], and Bufanda [Scarf] ($a \approx 2.6-2.7$), while the lowest value corresponded to Comedor ($a=1.97$).

These findings illustrate how certain items are particularly sensitive to differences in ability, while others are more easily accessed regardless of ability level. Fig 2 (left panel: Trial 4; right panel: Recognition) displays the corresponding Item Characteristic Curves (ICCs), highlighting the sharper slopes in Recognition that reflect stronger discrimination. The complete set of parameter estimates for all trials is provided in S2 Appendix. Based on these parameter estimates, ability scores (θ) were calculated for each participant in each of the six trials, which served as the outcome variables in the subsequent analyses examining demographic effects.

Demographic variables effect

The influence of demographic covariates on children’s performance across trials was examined using a multivariate framework. The initial specification of the linear mixed-effects regression model included age, age², MPE, sex, country, and all their second-order interactions. After variable selection process, the final linear mixed-effects regression revealed several significant interactions influencing ability scores (see Table 3). A robust effect was observed for the interaction between $\ln(\text{MPE})$ and country (Fig 3A): although all countries showed increasing ability scores with higher MPE, participants from Ecuador consistently achieved higher performance than those from the other countries. A second important effect was the age \times trial interaction (Fig 3B): in Free Recall–Trial 1, ability scores showed minimal improvement with age, suggesting that

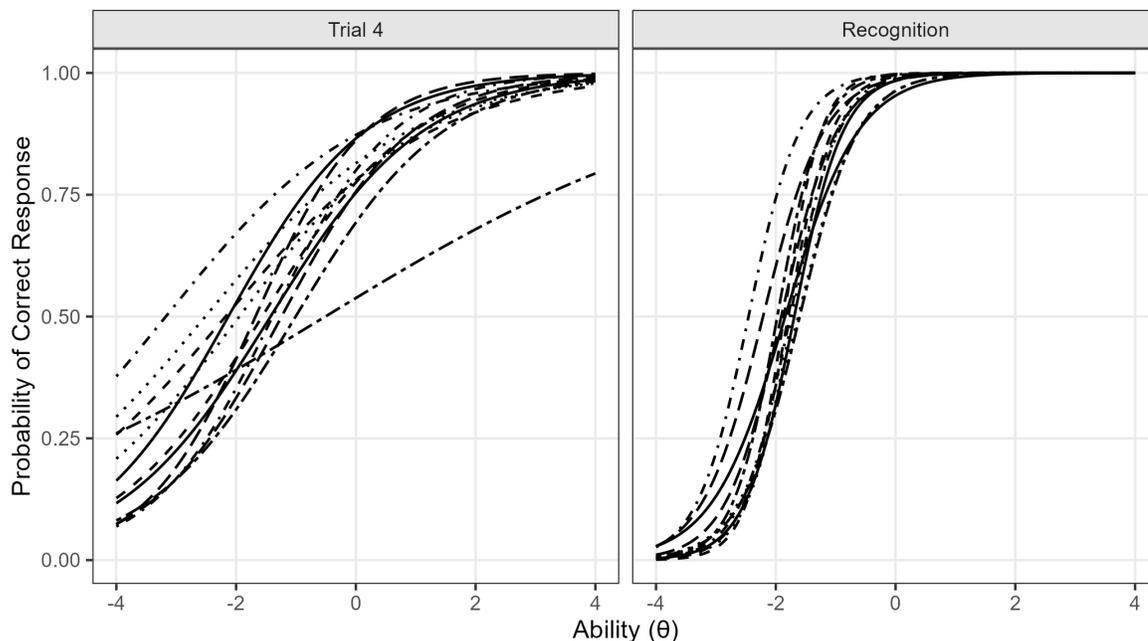


Fig 2. Item characteristic curves for Trial 4 and Recognition. Item characteristic curves (ICCs) from the two-parameter logistic (2PL) model for Trial 4 (left panel) and Recognition (right panel). The x-axis represents the latent ability level (θ), with higher values indicating better performance, whereas the y-axis indicates the probability of correctly responding to a given item. Each curve corresponds to one of the 12 items, and its horizontal position reflects item difficulty (b), while the steepness of the slope reflects item discrimination (a). Compared with Trial 4, the ICCs in the Recognition condition are notably steeper, indicating higher discrimination values and thus greater sensitivity in differentiating among individuals across ability levels.

<https://doi.org/10.1371/journal.pone.0341237.g002>

Table 3. Final linear mixed model.

Predictor	Beta	Standard Error	t	p-value
Intercept	-0.302	0.142	-2.123	0.034
Age	6.515	1.660	3.818	<0.001
Age ²	0.923	1.657	0.557	0.573
ln(MPE + 1)	0.118	0.054	2.189	0.029
Trial 2	-0.154	0.039	-3.977	<0.001
Trial 3	-0.142	0.039	-3.654	<0.001
Trial 4	-0.117	0.039	-3.007	0.003
Trial 5	-0.122	0.039	-3.142	0.002
Trial 6	-0.095	0.039	-2.452	0.014
Sex (Girl)	0.023	0.041	0.567	0.573
Country (Ecuador)	-0.020	0.199	-0.098	0.919
Country (Honduras)	-0.086	0.251	-0.344	0.733
Country (Spain)	-0.265	0.279	-0.674	0.342
Age*Trial 2	13.909	1.982	7.017	<0.001
Age ² *Trial 2	-7.943	1.980	-4.011	<0.001
Age*Trial 3	18.500	1.982	9.333	<0.001
Age ² *Trial 3	-9.077	1.980	-4.583	<0.001
Age*Trial 4	15.631	1.982	7.886	<0.001
Age ² *Trial 4	-11.924	1.980	-6.021	<0.001
Age*Trial 5	17.740	1.982	8.950	<0.001
Age ² *Trial 5	-13.021	1.980	-6.575	<0.001
Age*Trial 6	14.274	1.982	7.201	<0.001
Age ² *Trial 6	-15.345	1.980	-7.748	<0.001
ln(MPE)*Country (Ecuador)	-0.016	0.074	-0.212	0.834
ln(MPE)*Country (Honduras)	0.033	0.094	0.354	0.725
ln(MPE)*Country (Spain)	0.045	0.101	0.440	0.474
Sex (Girl)*Country (Ecuador)	0.100	0.055	1.820	0.069
Sex (Girl)*Country (Honduras)	-0.039	0.066	-0.596	0.551
Sex (Girl)*Country (Spain)	0.027	0.061	0.375	0.655
Trial 2*Country (Ecuador)	0.156	0.052	3.010	0.003
Trial 3*Country (Ecuador)	0.089	0.052	1.725	0.084
Trial 4*Country (Ecuador)	0.034	0.052	0.658	0.510
Trial 5*Country (Ecuador)	0.072	0.052	1.382	0.166
Trial 6*Country (Ecuador)	0.040	0.052	0.781	0.434
Trial 2*Country (Honduras)	0.046	0.062	0.749	0.454
Trial 3*Country (Honduras)	0.132	0.062	2.137	0.032
Trial 4*Country (Honduras)	0.117	0.062	1.897	0.058
Trial 5*Country Honduras)	0.119	0.062	1.925	0.054
Trial 6*Country (Honduras)	0.104	0.062	1.689	0.091
Trial 2*Country (Spain)	0.420	0.057	7.309	<0.001
Trial 3*Country (Spain)	0.398	0.057	6.933	<0.001
Trial 4*Country (Spain)	0.381	0.057	6.640	<0.001
Trial 5*Country (Spain)	0.346	0.057	6.027	<0.001
Trial 6*Country (Spain)	0.276	0.057	4.808	<0.001

Note: MPE = Mean Parent years of Education; $\sigma_{\epsilon} = 0.571$. This table shows the variables that influence children's performance on TAMV-I.

<https://doi.org/10.1371/journal.pone.0341237.t003>

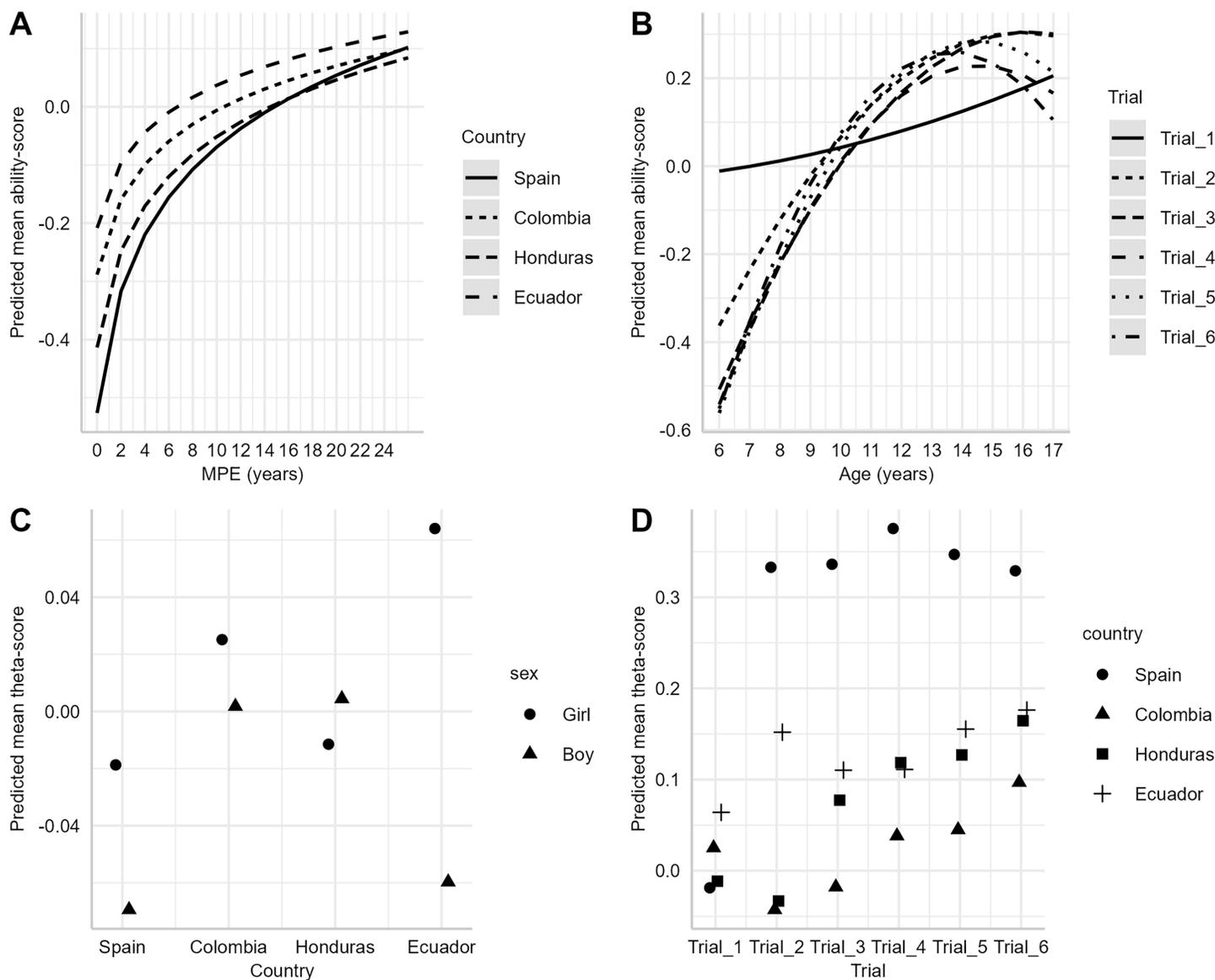


Fig 3. Final Multiple Linear Regression Model. Mixed model predictions for theta scores according to (a) years of parental education (MPE), (b) student age, (c) country and sex, and (d) trial and country. The lines represent the mean predicted ability score (theta); a consistent main effect of country is observed (Spain and Colombia outperforming Honduras and Ecuador), while age and trial show smaller differences, and for sex by country where ecuatorian girl has better performance than boys.

<https://doi.org/10.1371/journal.pone.0341237.g003>

age exerted little influence in the initial trial. In contrast, in subsequent trials performance increased until around 13 years of age, after which it declined. A third relevant effect was the sex × country interaction (Fig 3C), which indicated that girls consistently outperformed boys across countries, with the exception of Honduras, and with more pronounced differences observed in Spain and Ecuador. Finally, the trial × country interaction (Fig 3D) revealed that children from Spain generally performed better in all trials, except in Free Recall–Trial 1 where children from Ecuador obtained the highest ability scores.

These results provided the foundation for developing normative data adjusted for each child’s demographic background. In practical terms, this means that the ability score obtained by a participant can be directly compared with that

of peers of the same age, sex, country, and MPE. Such adjustments are essential to avoid misleading conclusions, for example, attributing a low score to cognitive difficulties when it may instead reflect differences in age or educational environment. The final normative data derived from these models therefore allow clinicians to interpret an individual child's performance more accurately and fairly within the appropriate reference group.

Normative data application

As an illustrative case, we considered a 17-year-old girl from Spain whose parents had an average of 18 years of education. Her performance was examined in Trial 6 (Recognition), where she scored 1 (correct) on all recognition items except for Item 1, where she scored 0. Based on this response pattern, and using the 2PL model (including item difficulty and discrimination parameters), the individual ability estimate was $\theta_i = -0.274$. To estimate the normative data for this participant, the procedure described in the Methods section was followed. First, the expected ability score for a participant with the same demographic profile was obtained from the final linear mixed-effects regression model (Table 3), resulting in $\hat{\theta}_i = 0.311$. Second the observed ability derived from the 2PL model ($\theta_i = -0.274$) was compared against the expected score predicted by the final linear mixed-effects model ($\hat{\theta}_i = 0.311$). Using the model's residual standard deviation ($\sigma_\varepsilon = 0.571$), the cumulative probability of obtaining an ability score less than or equal to the observed value was calculated as 0.153. Finally, this probability was multiplied by 100 to provide the normative percentile, corresponding to the 15.3th percentile. In other words, the participant's performance was higher than approximately 15% of peers matched on age, sex, country, and parental education.

Given its complexity, an online calculator has been developed to facilitate clinical practice which is based on the platform <https://www.rstudio.com/products/shiny/>. It allows for the computation of ability (θ) scores, as well as demographically adjusted z-scores and percentiles. Clinical psychologists only need to input specific patient information requested by the calculator, including item-by-item test responses (1 = correct; 0 = incorrect), age, MPE, country, and sex. This tool is accessible free of charge to all users at https://diegorivera.shinyapps.io/calculator_tamvi_tri/

Discussion

The objectives of this study were threefold: (1) To evaluate the discriminative ability and difficulty of TAMV-I items across different trials, analyzing the informative contribution of items on ability levels as a function of parameters obtained from the 2PL model, (2) To develop normative data for the TAMV-I using IRT models, and LMM, and (3) To provide a practical and accessible tool for professionals to calculate ability scores, zeta scores and adjusted percentiles for the TAMV-I to facilitate their clinical practice.

Our results confirm that the 2PL model offered a superior fit compared to the Rasch model, supporting the use of IRT as the methodological foundation for TAMV-I normative data. This empirical evidence is consistent with the broader limitations of Classical Test Theory (CTT), which assumes constant measurement error across individuals, relies on univariate adjustments for each score, and does not incorporate item-level parameters. Such limitations can inflate error rates, obscure inter-trial dependencies, and yield parameter estimates that are sample-dependent and potentially biased [46]. In contrast, IRT models, including Rasch and 2PL, explicitly account for item characteristics, but the 2PL model provides greater flexibility by estimating both item difficulty and discrimination, which resulted in a better fit for our data [47]. This framework provides richer psychometric information, supports the development of cross-national and covariate-adjusted norms, and facilitates adaptive testing designs. The analysis of item discrimination and difficulty parameters further confirmed this point, revealing substantial variability among items and underscoring the importance of an item-level approach. For instance, in Free recall – Trial 1, the item Zapato [Shoe] was the best item for discrimination, suggesting its capacity to differentiate between individuals with varying levels of cognitive ability. Conversely, Nariz [Nose] exhibited a negative discrimination value, indicating a reduced efficiency in distinguishing between different ability levels. The difficulty parameters also showed a similar pattern. Sillón [Armchair] presented an exceptionally high difficulty value in Free recall – Trial

1, becoming a highly challenging item. In contrast, Bufanda [Scarf] showed a very low difficulty value, and, therefore, extremely easy item.

Some items showed notable variations in their discrimination and difficulty parameters. Zapato [Shoe] exhibited consistent moderate to high discrimination values, indicating its reliability in distinguishing between different levels of cognitive ability across trials. Sillón [Armchair] showed significant fluctuations in difficulty, ranging from highly difficult in Free recall – Trial 1 to progressively easier in subsequent trials and Delayed Recall. Nevertheless, it remained the most difficult item in each trial. Nariz [Nose] consistently had negative or low discrimination values, as well as low difficulties values, suggesting its poor capacity to differentiate between individuals because it is an easy item.

Such disparities were consistently evident across subsequent trials, with items like Armario [Wardrobe] in Free recall – Trial 2 demonstrating high discrimination and items like Escritorio [Desk] in the same trial exhibiting extreme ease. Although the words for the TAMV-I were selected based on word frequency in Spanish, calculated from both Spanish and Cuban samples [53], the observed variability in item discrimination and difficulty could be attributed to differences in word frequency and familiarity for the children participating in this study.

Differences were also observed between Free recall trials, Delayed recall, and Recognition tasks, and interestingly, as trials progress, the scores showed better parameters. This may reflect the learning process of the person. This finding aligns with the results reported by [62] in their research on the California Verbal Learning Test – Second Edition (CVLT-II). Also, Free recall trials, particularly the earlier ones, might be influenced by learning effects and potential fatigue, as evidenced by the changing discrimination and difficulty values of items like Zapato [Shoe] and Sillón [Armchair]. The increase in discrimination values for Zapato [Shoe] from Free recall – Trial 1 ($a=0.60$) to Trial 3 ($a=1.39$) suggests that participants improved their ability to recall this item with repeated exposure. In delayed recall, on the other hand, items such as Blusa [Blouse] maintained high discrimination ($a=1.24$), indicating effective differentiation even after a delay, while Nariz [Nose] showed low discrimination ($a=0.57$), highlighting its reduced efficiency in delayed contexts. This balanced mix of item difficulties and discrimination ability is essential to effectively span the spectrum of cognitive abilities being measured.

Analysis of the effect of demographic variables using LMM revealed significant interactions showing the effect of factors such as age, MPE, sex, and country of origin on the learning and memory skills assessed. As observed in Van der Elst's [45], analysis of the Rey Auditory Verbal Learning Test (RAVLT), the age of participants influenced performance across all trials. This interaction showed expected patterns associated with maturation processes of the brain [63]. Interestingly, for Free recall – Trial 1, the required skill level was high and performance increased slowly with age. This could be because participants faced first trial without the familiarity or practice acquired in later trials, suggesting that this trial may be novel, complex, and measuring a different cognitive domain than the other trials, primarily attentional ability. In the later trials, while this trend in performance was also observed, at 13 years there is a plateauing and subsequent decline, suggesting a typical cognitive developmental curve. These findings are in line with previous literature that identifies the peak of cognitive development in early adolescence followed by a stabilization or mild decline [64,65].

The interaction between MPE and country showed that, in general, as expected, higher parental years of education is associated with better performance on the TAMV-I. Higher MPE is usually associated with a more cognitively stimulating family environment, which could facilitate the development of learning and memory skills as suggested by multiple studies [66–68]. The outstanding improvement in children from Ecuador with high levels of MPE (see Fig 2), suggest that, in this country, the benefits of high parental education may be more pronounced than in the rest. A possible explanation for these differences could be that parental stimulation at home, informed by parental education, is more impactful in Ecuador due to a less effective educational system influenced by various socioeconomic and cultural factors within the country [69].

The sex by country interaction reflected better performance of girls compared to boys in almost all countries except Honduras. This finding is consistent with previous research indicating that girls tend to score higher on tests that assess verbal and memory skills than boys [70,71]. The most pronounced differences, observed in Spain and Ecuador, could be influenced by cultural factors, such as gender expectations and/or parenting styles that favor the development of verbal

skills in girls [72]. At the other hand in the case of Honduras, women face several sociocultural disadvantages that impact their performance, these challenges include a higher rate of illiteracy and the traditional expectation that their duties are primarily focused on domestic tasks [73].

Regarding the trial by country interaction, Spanish children performed better, with the exception of the first trial of the test, where children from Ecuador outperformed children from other countries. This result suggests that there may be differences in task preparation and familiarity between countries. The higher scores of Spanish children on trials 2–6 could be due to greater exposure to educational practices that emphasize learning and memory skills from an early age. On the other hand, the higher performance of Ecuadorian children in the first trial could indicate differences in motivation or initial approach to tasks.

This paper presents with several strengths. Firstly, the analyses were performed on a large sample from various countries in Latin America and Spain, enhancing both representativeness and generalizability. Secondly, the study employed a hybrid approach combining IRT and continuous-norming techniques, which allowed for greater precision by leveraging IRT-derived ability scores and adjusting for demographic factors, including country of origin, via regression analysis. Thirdly, the use of regression-based norms allowed us to control for demographic variables related to cognitive performance, and therefore the normative data produced is applicable to populations with demographic differences captured in the regression equation. To the authors' knowledge, the methods applied in this study have been rarely used despite its benefits. Additionally, an accessible online calculator is provided at https://diegorivera.shinyapps.io/calculator_tamvi_tri/.

However, the study also has its limitations. While linear and quadratic models were tested, other polynomial models (such as cubic or logarithmic functions) were not explored, which might have improved model fit, but deviated from the principle of parsimony. Moreover, the study could have included additional variables, like socioeconomic status, quality of education. These aspects could be considered in future research, although most normative data studies use sex, age, and education because these variables are more easily standardized across different test administrations and populations. In addition, while the normative data provided here are robust for the populations sampled in each country, they should not be assumed to capture the full variability of all subgroups. For instance, ethnic and racial minorities may present distinct cultural or educational experiences that can significantly impact their performance on certain assessments. Therefore, clinicians are advised to apply the norms with caution in such subpopulations, as relying on generalized data may reduce diagnostic accuracy for these specific groups [74].

The study's findings have significant clinical implications. Normative data in Latin America are scarce compared to the extensive data available in the United State of America and European countries. Enhancing normative tools for under-represented populations is likely to advance neuropsychological practice by providing more appropriate reference populations. Additionally, using a distribution of theta scores instead of true scores can increase precision, thereby improving diagnosis and treatment. Considering and controlling for the impact of demographic variables when deriving scores will also enhance precision, especially in Latin American countries with significant demographic disparities.

Conclusion

The present study provides robust normative data for the TAMV-I using IRT and LMM models. The choice of the 2PL model, based on the BIC fit indices, over the Rasch model, demonstrates that this model allows a superior fit for the test data. In addition, the findings obtained reflect the importance of considering demographic and contextual variables in the interpretation of the results. Parental education, age, sex and country of origin were shown to have a significant influence on the scores obtained by the participants. Finally, the online tool and normative data developed in this study represent a valuable contribution to clinical practice, facilitating accurate and accessible score calculations for psychologists.

Supporting information

S1 File. Anonymized dataset.
(CSV)

Author contributions

Conceptualization: Laiene Olabarrieta-Landa, Erick Orozco-Acosta, Diego Rivera.

Data curation: Eliana María Fuentes Mendoza, Esperanza Vergara-Moragues, Natalia Albaladejo-Blázquez, Natalia Cadavid-Ruiz, María José Irias Escher.

Formal analysis: Eliana María Fuentes Mendoza, Erick Orozco-Acosta, Diego Rivera.

Investigation: Alberto Rodríguez-Lorenzana, Guido Mascialino.

Methodology: Diego Rivera.

Project administration: Juan Carlos Arango-Lasprilla.

Supervision: Laiene Olabarrieta-Landa, Juan Carlos Arango-Lasprilla, Diego Rivera.

Visualization: Erick Orozco-Acosta.

Writing – original draft: Eliana María Fuentes Mendoza, Laiene Olabarrieta-Landa, Alberto Rodríguez-Lorenzana, Guido Mascialino, Carlos José de los Reyes-Aragón, Diego Rivera.

Writing – review & editing: Laiene Olabarrieta-Landa, Alberto Rodríguez-Lorenzana, Guido Mascialino, Juan Carlos Arango-Lasprilla, Erick Orozco-Acosta, Diego Rivera.

References

1. Cohen LE, Waite-Stupiansky S. Theories of Early Childhood Education: Developmental, Behaviorist, and Critical. 2nd ed. New York: Routledge; 2022. <https://doi.org/10.4324/9781003288077>
2. De los Reyes-Aragon CJ, Amar Amar J, De Castro Correa A, Lewis Harb S, Madariaga C, Abello-Llanos R. The Care and development of children living in contexts of poverty. *J Child Fam Stud*. 2016;25(12):3637–43. <https://doi.org/10.1007/s10826-016-0514-6>
3. Tudge JRH, Merçon-Vargas EA, Liang Y, Payir A. The Importance of Urie Bronfenbrenner's Bioecological Theory for Early Childhood Education. 2nd ed. Routledge; 2022.
4. Berk L. Child development. Pearson Higher Education AU; 2015.
5. Navarro JL, Tudge JRH. Technologizing Bronfenbrenner: Neo-ecological Theory. *Curr Psychol*. 2022;:1–17. <https://doi.org/10.1007/s12144-022-02738-3> PMID: 35095241
6. Ramos AA, Hamdan AC, Machado L. A meta-analysis on verbal working memory in children and adolescents with ADHD. *Clin Neuropsychol*. 2020;34(5):873–98. <https://doi.org/10.1080/13854046.2019.1604998> PMID: 31007130
7. Skodzik T, Holling H, Pedersen A. Long-term memory performance in adult ADHD. *J Atten Disord*. 2017;21(4):267–83. <https://doi.org/10.1177/1087054713510561> PMID: 24232170
8. Hedenius M, Lum JAG, Bölte S. Alterations of procedural memory consolidation in children with developmental dyslexia. *Neuropsychology*. 2021;35(2):185–96. <https://doi.org/10.1037/neu0000708> PMID: 33211512
9. McCloskey M, Rapp B. Developmental dysgraphia: an overview and framework for research. *Cogn Neuropsychol*. 2017;34(3–4):65–82. <https://doi.org/10.1080/02643294.2017.1369016> PMID: 28906176
10. Luoni C, Scorza M, Stefanelli S, Fagiolini B, Termine C. A neuropsychological profile of developmental dyscalculia: the role of comorbidity. *J Learn Disabil*. 2023;56(4):310–23. <https://doi.org/10.1177/00222194221102925> PMID: 35726739
11. Hronis A, Roberts L, Kneebone II. A review of cognitive impairments in children with intellectual disabilities: implications for cognitive behaviour therapy. *Br J Clin Psychol*. 2017;56(2):189–207. <https://doi.org/10.1111/bjc.12133> PMID: 28397306
12. Vicari S, Costanzo F, Menghini D. Chapter Four - Memory and Learning in Intellectual Disability. In: Hodapp RM, Fidler DJ, editors. *International Review of Research in Developmental Disabilities*. Academic Press; 2016. pp. 119–148. <https://doi.org/10.1016/bs.irrdd.2016.05.003>
13. Boucher J, Anns S. Memory, learning and language in autism spectrum disorder. *Autism Dev Lang Impair*. 2018;3. <https://doi.org/10.1177/2396941517742078>
14. Desautay P, Briant AR, Bowler DM, Ring M, Gérardin P, Baleyte J-M, et al. Memory in autism spectrum disorder: a meta-analysis of experimental studies. *Psychol Bull*. 2020;146(5):377–410. <https://doi.org/10.1037/bul0000225> PMID: 32191044
15. Prabhakar J, Coughlin C, Ghetti S. The neurocognitive development of episodic prospection and its implications for academic achievement. *Mind Brain Educ*. 2016;10(3):196–206. <https://doi.org/10.1111/mbe.12124>
16. Simone AN, Marks DJ, Bédard A-C, Halperin JM. Low working memory rather than ADHD symptoms predicts poor academic achievement in school-aged children. *J Abnorm Child Psychol*. 2018;46(2):277–90. <https://doi.org/10.1007/s10802-017-0288-3> PMID: 28357519

17. Sjöwall D, Bohlin G, Rydell A-M, Thorell LB. Neuropsychological deficits in preschool as predictors of ADHD symptoms and academic achievement in late adolescence. *Child Neuropsychol*. 2017;23(1):111–28. <https://doi.org/10.1080/09297049.2015.1063595> PMID: [26212755](https://pubmed.ncbi.nlm.nih.gov/26212755/)
18. Stipek D, Valentino RA. Early childhood memory and attention as predictors of academic growth trajectories. *J Educ Psychol*. 2015;107(3):771–88. <https://doi.org/10.1037/edu0000004>
19. Bullard CC, Alderson RM, Roberts DK, Tatsuki MO, Sullivan MA, Kofler MJ. Social functioning in children with ADHD: an examination of inhibition, self-control, and working memory as potential mediators. *Child Neuropsychol*. 2024;30(7):987–1009. <https://doi.org/10.1080/09297049.2024.2304375> PMID: [38269494](https://pubmed.ncbi.nlm.nih.gov/38269494/)
20. McQuade JD, Murray-Close D, Shoulberg EK, Hoza B. Working memory and social functioning in children. *J Exp Child Psychol*. 2013;115(3):422–35. <https://doi.org/10.1016/j.jecp.2013.03.002> PMID: [23665178](https://pubmed.ncbi.nlm.nih.gov/23665178/)
21. Brylka M, Cygan HB. Selective short-term memory impairment for verbalizable visual objects in children with developmental language disorder. *Res Dev Disabil*. 2024;144:104637. <https://doi.org/10.1016/j.ridd.2023.104637> PMID: [38035638](https://pubmed.ncbi.nlm.nih.gov/38035638/)
22. Peng P, Fuchs D. A meta-analysis of working memory deficits in children with learning difficulties: is there a difference between verbal domain and numerical domain? *J Learn Disabil*. 2016;49(1):3–20. <https://doi.org/10.1177/0022219414521667> PMID: [24548914](https://pubmed.ncbi.nlm.nih.gov/24548914/)
23. Signori VDA, Watanabe TM, de Pereira APA. Prospective memory instruments for the assessment of children and adolescents: a systematic review. *Psicol Reflex Crit*. 2024;37(1):17. <https://doi.org/10.1186/s41155-024-00300-7> PMID: [38709384](https://pubmed.ncbi.nlm.nih.gov/38709384/)
24. Arango-Lasprilla JC, Stevens L, Morlett Paredes A, Ardila A, Rivera D. Profession of neuropsychology in Latin America. *Appl Neuropsychol Adult*. 2017;24(4):318–30. <https://doi.org/10.1080/23279095.2016.1185423> PMID: [27282450](https://pubmed.ncbi.nlm.nih.gov/27282450/)
25. Kasperek A, Kingma A, de Aguiar V. The 10-word auditory verbal learning test and vocabulary performance in 4- and 5-year-old children. *J Speech Lang Hear Res*. 2023;66(11):4464–80. https://doi.org/10.1044/2023_JSLHR-22-00706 PMID: [37774742](https://pubmed.ncbi.nlm.nih.gov/37774742/)
26. Oliveira RM, Mograbi DC, Gabrig IA, Charchat-Fichman H. Normative data and evidence of validity for the Rey auditory verbal learning test, verbal fluency test, and stroop test with Brazilian children. *Psychol Neurosci*. 2016;9(1):54–67. <https://doi.org/10.1037/pne0000041>
27. Verroulx K, Hirst RB, Lin G, Peery S. Embedded performance validity indicator for children: California Verbal Learning Test - Children's Edition, forced choice. *Appl Neuropsychol Child*. 2019;8(3):206–12. <https://doi.org/10.1080/21622965.2018.1426463> PMID: [29412011](https://pubmed.ncbi.nlm.nih.gov/29412011/)
28. Bezdíček O, Stepankova H, Moták L, Axelrod BN, Woodard JL, Preiss M, et al. Czech version of Rey Auditory Verbal Learning test: normative data. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn*. 2014;21(6):693–721. <https://doi.org/10.1080/13825585.2013.865699> PMID: [24344673](https://pubmed.ncbi.nlm.nih.gov/24344673/)
29. Poreh A, Teaford M. Normative data and construct validation for a novel nonverbal memory test. *Arch Assess Psychol*. 2017;7:43–60.
30. Rodríguez-Cancino M, Vizcarra MB, Concha-Salgado A. Propiedades Psicométricas de la Escala WISC-V en Escolares Rurales Chilenos. *Psykhe (Santiago)*. 2022;31(2). <https://doi.org/10.7764/psykhe.2020.22529>
31. Rosselli-Cock M, Matute E, Ardila A, Botero-Gómez VE, Tangarife-Salazar GA, Echevarría-Pulido SE. Neuropsychological assessment of children: a test battery for children between 5 and 16 years of age. A Colombian normative study. *RN*. 2004;:720–31.
32. Schnurbusch Gallardo CS, Suárez Yepes N, Ortiz Tejera D, de los Reyes Aragón CJ. Datos normativos para la batería de evaluación neuropsicológica de lectura, escritura y funciones cognitivas (ENLEF). *Psicología desde el Caribe: revista del Programa de Psicología de la Universidad del Norte*. 2018;35:252–67.
33. Casaletto KB, Heaton RK. Neuropsychological assessment: past and future. *J Int Neuropsychol Soc*. 2017;23(9–10):778–90. <https://doi.org/10.1017/S1355617717001060> PMID: [29198281](https://pubmed.ncbi.nlm.nih.gov/29198281/)
34. Arango-Lasprilla JC. Commonly used Neuropsychological Tests for Spanish Speakers: Normative Data from Latin America. *NeuroRehabilitation*. 2015;37(4):489–91. <https://doi.org/10.3233/NRE-151276> PMID: [26577888](https://pubmed.ncbi.nlm.nih.gov/26577888/)
35. Kabuba N, Anitha Menon J, Franklin DR Jr, Heaton RK, Hestad KA. Use of Western Neuropsychological Test Battery in Detecting HIV-Associated Neurocognitive Disorders (HAND) in Zambia. *AIDS Behav*. 2017;21(6):1717–27. <https://doi.org/10.1007/s10461-016-1443-5> PMID: [27278547](https://pubmed.ncbi.nlm.nih.gov/27278547/)
36. Malda M, van de Vijver FJR, Srinivasan K, Transler C, Sukumar P. Traveling with cognitive tests: testing the validity of a KABC-II adaptation in India. *Assessment*. 2010;17(1):107–15. <https://doi.org/10.1177/1073191109341445> PMID: [19745212](https://pubmed.ncbi.nlm.nih.gov/19745212/)
37. Fasfous AF, Al-Joudi HF, Puente AE, Pérez-García M. Neuropsychological measures in the Arab World: a systematic review. *Neuropsychol Rev*. 2017;27(2):158–73. <https://doi.org/10.1007/s11065-017-9347-3> PMID: [28624899](https://pubmed.ncbi.nlm.nih.gov/28624899/)
38. Ben-David BM, Erel H, Goy H, Schneider BA. “Older is always better”: age-related differences in vocabulary scores across 16 years. *Psychol Aging*. 2015;30(4):856–62. <https://doi.org/10.1037/pag0000051> PMID: [26652725](https://pubmed.ncbi.nlm.nih.gov/26652725/)
39. Farkas G, Beron K. The detailed age trajectory of oral vocabulary knowledge: differences by class and race. *Soc Sci Res*. 2004;33(3):464–97. <https://doi.org/10.1016/j.ssresearch.2003.08.001>
40. Riva A, Musetti A, Bombà M, Milani L, Montrasi V, Nacinovich R. Language-related skills in Bilingual children with specific learning disorders. *Front Psychol*. 2021;11:564047. <https://doi.org/10.3389/fpsyg.2020.564047> PMID: [33551894](https://pubmed.ncbi.nlm.nih.gov/33551894/)
41. Rivera D, Arango-Lasprilla JC. Methodology for the development of normative data for Spanish-speaking pediatric populations. *NeuroRehabilitation*. 2017;41(3):581–92. <https://doi.org/10.3233/NRE-172275> PMID: [29036850](https://pubmed.ncbi.nlm.nih.gov/29036850/)
42. Matute E, Rosselli M, Ardila A, Ostrosky Solís F. Evaluación neuropsicológica infantil (ENI-2). Segunda edición. Manual Moderno; 2013.

43. Rivera D, Forte A, Olabarrieta-Landa L, Perrin PB, Arango-Lasprilla JC. Methodology for the generation of normative data for the U.S. adult Spanish-speaking population: a Bayesian approach. *NeuroRehabilitation*. 2024;55(2):155–67. <https://doi.org/10.3233/NRE-240149> PMID: [39302390](https://pubmed.ncbi.nlm.nih.gov/39302390/)
44. Rivera D, Olabarrieta-Landa L, Rabago Barajas BV, Iriás Escher MJ, Saracostti Schwartzman M, Ferrer-Cascales R, et al. Newly developed Learning and Verbal Memory Test (TAMV-I): Normative data for Spanish-speaking pediatric population. *NeuroRehabilitation*. 2017;41(3):695–706. <https://doi.org/10.3233/NRE-172249> PMID: [29036849](https://pubmed.ncbi.nlm.nih.gov/29036849/)
45. Van der Elst W, Molenberghs G, van Tetering M, Jolles J. Establishing normative data for multi-trial memory tests: the multivariate regression-based approach. *Clin Neuropsychol*. 2017;31(6–7):1173–87. <https://doi.org/10.1080/13854046.2017.1294202> PMID: [28276864](https://pubmed.ncbi.nlm.nih.gov/28276864/)
46. Zanon C, Hutz CS, Yoo H, Hambleton RK. An application of item response theory to psychological test development. *Psicol Refl Crít*. 2016;29(1). <https://doi.org/10.1186/s41155-016-0040-x>
47. Jabrayilov R, Emons WHM, Sijtsma K. Comparison of Classical Test Theory and Item Response Theory in Individual Change Assessment. *Appl Psychol Meas*. 2016;40(8):559–72. <https://doi.org/10.1177/0146621616664046> PMID: [29881070](https://pubmed.ncbi.nlm.nih.gov/29881070/)
48. West BT, Welch KB, Galecki AT. *Linear Mixed Models: A Practical Guide Using Statistical Software*. 2nd ed. New York: Chapman and Hall/CRC; 2014. <https://doi.org/10.1201/b17198>
49. Innocenti F, Tan FES, Candel MJJM, van Breukelen GJP. Sample size calculation and optimal design for regression-based norming of tests and questionnaires. *Psychol Methods*. 2023;28(1):89–106. <https://doi.org/10.1037/met0000394> PMID: [34383531](https://pubmed.ncbi.nlm.nih.gov/34383531/)
50. Brown L, Sherbenou RJ, Johnsen SK. *Test de inteligencia no verbal: TONI-2*. TEA Ediciones. 2009.
51. Kovacs M. *Children's Depression Inventory*. Manual/Multi-Health Systems Inc.; 1992.
52. Bush K, Kivlahan DR, McDonnell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *Arch Intern Med*. 1998;158(16):1789–95. <https://doi.org/10.1001/archinte.158.16.1789> PMID: [9738608](https://pubmed.ncbi.nlm.nih.gov/9738608/)
53. Rivera D, Olabarrieta-Landa L, Arango-Lasprilla JC. Diseño y creación del Test de Aprendizaje y Memoria Verbal Infantil (TAMV-I) en población hispano hablante de 6 a 17 años de edad. In: Arango-Lasprilla JC, Rivera D, Olabarrieta-Landa L, editors. *Neuropsicología infantil*. Bogotá: Manual Moderno; 2017. pp. 316–38.
54. Rizopoulos D. ltm: AnRPackage for latent variable modeling and item response theory analyses. *J Stat Soft*. 2006;17(5). <https://doi.org/10.18637/jss.v017.i05>
55. Faraway JJ. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Chapman and Hall/CRC; 2016.
56. Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest Package: tests in linear mixed effects models. *J Stat Soft*. 2017;82(13). <https://doi.org/10.18637/jss.v082.i13>
57. Patterson HD, Thompson R. Recovery of inter-block information when block sizes are unequal. *Biometrika*. 1971;58(3):545–54. <https://doi.org/10.1093/biomet/58.3.545>
58. Kramlinger P, Schneider U, Krivobokova T. Uniformly valid inference based on the Lasso in linear mixed models. *J Multiv Anal*. 2023;198:105230. <https://doi.org/10.1016/j.jmva.2023.105230>
59. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning: with Applications in R*. New York, NY: Springer US; 2021. <https://doi.org/10.1007/978-1-0716-1418-1>
60. R Core Team. *The R Stats Package*. 2021. Available from: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>
61. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models Using lme4. *J Stat Soft*. 2015;67(1). <https://doi.org/10.18637/jss.v067.i01>
62. Thiruselvam I, Hoelzle JB. Refined Measurement of Verbal Learning and Memory: Application of Item Response Theory to California Verbal Learning Test - Second Edition (CVLT-II) Learning Trials. *Arch Clin Neuropsychol*. 2020;35(1):90–104. <https://doi.org/10.1093/arclin/acy097> PMID: [30615062](https://pubmed.ncbi.nlm.nih.gov/30615062/)
63. Semrud-Clikeman M. *Research in brain function and learning*. 2010 [cited 11 Jan 2025]. Available from: <https://www.apa.org/education-career/k12/brain-function>
64. Crone EA, Dahl RE. Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nat Rev Neurosci*. 2012;13(9):636–50. <https://doi.org/10.1038/nrn3313>
65. Gogtay N, Giedd JN, Lusk L, Hayashi KM, Greenstein D, Vaituzis AC, et al. Dynamic mapping of human cortical development during childhood through early adulthood. *Proc Natl Acad Sci U S A*. 2004;101(21):8174–9. <https://doi.org/10.1073/pnas.0402680101> PMID: [15148381](https://pubmed.ncbi.nlm.nih.gov/15148381/)
66. Kalil A, Ryan R, Corey M. Diverging destinies: maternal education and the developmental gradient in time with children. *Demography*. 2012;49(4):1361–83. <https://doi.org/10.1007/s13524-012-0129-5> PMID: [22886758](https://pubmed.ncbi.nlm.nih.gov/22886758/)
67. Magnuson K, Duncan GJ. Can early childhood interventions decrease inequality of economic opportunity? Russell Sage Foundation *J Soc Sci*. 2016;2(2):123–41. <https://doi.org/10.7758/RSF.2016.2.2.05> PMID: [30135867](https://pubmed.ncbi.nlm.nih.gov/30135867/)
68. Rosenzweig MR, Bennett EL. Psychobiology of plasticity: effects of training and experience on brain and behavior. *Behav Brain Res*. 1996;78(1):57–65. [https://doi.org/10.1016/0166-4328\(95\)00216-2](https://doi.org/10.1016/0166-4328(95)00216-2) PMID: [8793038](https://pubmed.ncbi.nlm.nih.gov/8793038/)

69. UNESCO. Global education monitoring report, 2020: Inclusion and education: all means all. UNESCO; 2020. Available from: <https://unesdoc.unesco.org/ark:/48223/pf0000373718>
70. Keith TZ, Reynolds MR, Patel PG, Ridley KP. Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock–Johnson III tests of cognitive abilities. *Intelligence*. 2008;36(6):502–25. <https://doi.org/10.1016/j.intell.2007.11.001>
71. Lowe PA, Mayfield JW, Reynolds CR. Gender differences in memory test performance among children and adolescents. *Arch Clin Neuropsychol*. 2003;18(8):865–78. <https://doi.org/10.1093/arclin/18.8.865>
72. Else-Quest NM, Hyde JS, Goldsmith HH, Van Hulle CA. Gender differences in temperament: a meta-analysis. *Psychol Bull*. 2006;132(1):33–72. <https://doi.org/10.1037/0033-2909.132.1.33> PMID: [16435957](https://pubmed.ncbi.nlm.nih.gov/16435957/)
73. Instituto Nacional de la Mujer. II Plan de Equidad e Igualdad de Género de Honduras 2010-2022. Instituto Nacional de la Mujer; 2010. Available from: https://oig.cepal.org/sites/default/files/honduras_2010_2022_piegh.pdf
74. Brickman AM, Cabo R, Manly JJ. Ethical issues in cross-cultural neuropsychology. *Appl Neuropsychol*. 2006;13(2):91–100. https://doi.org/10.1207/s15324826an1302_4 PMID: [17009882](https://pubmed.ncbi.nlm.nih.gov/17009882/)