ISSN 2055-0286
Advances in Social Sciences Research Journal; Vol. 13 No. 11 (2026) (491-506)
https://doi.org/10.14738/assrj.1311.12703

Measuring Processing Speed with Rasch Count Models: A Conway-Maxwell-Poisson Scale and Norms for Four Open Bafacalo Tests

Cristiano Mauro Assis Gomes & Jhonys de Araujo

Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil

Abstract: In speed tests with very easy items, the time limit constrains performance, and the score is a count of items solved. Rasch count models suit such scores but are rarely used. This study applies them to build a processing speed (Gs) measure from four open tests of the Battery of High-Order Cognitive Factors (BAFACALO): Perceptual Speed Tests 1, 2, and 3 (P1, P2, and P3) and the Numerical Speed Test (N), with 230 students from one Brazilian public high school. The Rasch Poisson Counts Model (RPCM) was tested first, since its total score applies directly to norms. It assumes a single underlying ability (unidimensionality) and a count variance equal to the mean (equidispersion). Eigenvalue analyses of scores and residuals supported unidimensionality. Equidispersion was violated in opposite directions: P1 varied more than the Poisson distribution allows, and P2 varied less. A Conway-Maxwell-Poisson (CMP) model with a dispersion parameter for each test fitted better than the RPCM and a single-dispersion CMP model, and its residuals supported both unidimensionality and the dispersion estimated for each test. The four scores can still be summed once each is divided by the dispersion of its test. This weighted sum ordered students almost exactly as the model (Spearman $\rho > .999$) and was converted into percentile ranks and T scores. The study thus provides validity evidence for a Gs measure from four open tests, with norms that users can apply without fitting the model. These norms remain preliminary until replicated in larger, more diverse samples.

Keywords: Processing speed, Rasch Poisson Counts Model, Conway-Maxwell-Poisson model, open tests, count data.

INTRODUCTION

The open science movement seeks to make research products accessible, transparent, and reusable, so that other researchers can examine, replicate, and extend them (Crüwell et al., 2019; Nosek et al., 2015). In psychology, these products include tests, on which much of the field's research depends.

Psychology can benefit from this movement through open tests. Anyone can use and adapt these tests free of charge, so they avoid the cost and licensing restrictions that limit the use of commercial tests in research and in low-resource settings. This openness also carries a risk, because a test with measurement problems or without validity evidence can compromise every study that applies it. Validity evidence for an open test reduces this risk and helps ensure that studies built on open tests rest on sound measures.

The Battery of High-Order Cognitive Factors (BAFACALO) is a battery of open tests. The Laboratory for Cognitive Architecture Mapping (LAICO) at the Federal University of Minas Gerais released the battery openly in 2021, first as preprints (Gomes & Nascimento, 2021a, 2021b, 2021c, 2021d, 2021e, 2021f, 2021g, 2021h, 2021i, 2021j, 2021k, 2021l,

2021m, 2021n; Gomes et al., 2021a, 2021b, 2021c, 2021d), and has since published several of the tests in peer-reviewed journals (Gomes & Araujo, 2025a, 2025b, 2025c, 2025d). The battery comprises 18 tests developed in 2006 on the basis of Carroll's (1993) Three-Stratum model and the Cattell-Horn-Carroll (CHC) model (Schneider & McGrew, 2018). It is designed to measure the general factor of intelligence, six broad factors, and 18 narrow factors.

The BAFACALO has accumulated validity evidence from several sources. The battery has factorial validity evidence (Gomes, 2010b, 2011, 2012; Gomes & Araujo, 2026a, 2026b; Gomes & Borges, 2009a, 2009b; Gomes, de Araújo, et al., 2014; Gomes et al., 2022; Gomes & Golino, 2015). It also has predictive (Gomes, 2010a, 2011; Gomes & Golino, 2012a; Gomes, Marques, & Golino, 2014), incremental (Alves et al., 2012; Gomes, 2012; Gomes et al., 2022; Gomes & Golino, 2012b), and convergent and divergent validity evidence (Gomes, 2010b, 2011, 2012; Gomes & Borges, 2009a, 2009b).

Four BAFACALO tests measure processing speed (Gs), namely Perceptual Speed Tests 1, 2, and 3 (P1, P2, and P3) and the Numerical Speed Test (N) (Gomes & Nascimento, 2021g, 2021i; Gomes et al., 2021c, 2021d). Previous studies have examined the factorial validity of these tests (Gomes, 2010b; X), but their format poses a challenge for psychometric measure construction. In all four tests, the items are simple enough that almost anyone could solve them given unlimited time, so the score is the number of items a person solves within the time limit. A psychometric measure can still be built from count scores of this kind, and the Rasch Poisson Counts Model (RPCM) is an appropriate option for this purpose. The RPCM takes the count itself as the datum, jointly models several tests of the same ability, and places persons and tests on a common scale (Baghaei & Doebler, 2019).

The present study uses the RPCM to build a measure of Gs from the scores on P1, P2, P3, and N. The study first evaluates whether the scores on the four tests violate the assumptions of the RPCM. If they do not, the four tests can be represented on a single scale of processing speed, and the study estimates the easiness parameter of each test on that scale and converts the resulting ability estimates into a score interpretation table referenced to the study sample. The scale would give users of the four tests a common measure of processing speed, and the table would give them a reference for interpreting the scores.

METHOD

Participants

Participants were drawn from the open dataset of Golino and Gomes (2014), which contains the scores of a convenience sample of 292 students from a public high school in Belo Horizonte, Brazil, on the 18 BAFACALO tests. The present study included only the 230 students with scores on all four processing speed tests (P1, P2, P3, and N). Of the participants, 122 (53.0%) were female, 107 (46.5%) male, and one (0.4%) did not report sex. Ages ranged from 14 to 20 years (M = 15.71, SD = 1.16), with 22 participants (9.6%) not reporting age. Participants attended the first (n = 121, 52.6%), second (n = 65, 28.3%), or third year of high school (n = 43, 18.7%), and one (0.4%) did not report grade. Household income was classified into three ranges in multiples of the 2006 Brazilian minimum wage (R$ 350, about USD 164 at R$ 2.1372 per US dollar on December 31, 2006). Of the

participants, 100 (43.5%) reported up to 5 minimum wages (about USD 819), 92 (40.0%) from 5 to 10 (about USD 1,638), and 36 (15.7%) above 10, and two (0.9%) did not report income.

Instruments

The score on each test is the number of correct responses within the time limit. The tests are intended for anyone who has at least begun lower secondary education, including adults who are no longer in school, and they can be administered to groups as well as to individuals. Perceptual Speed Test 1 (P1) is freely and openly available in Gomes and Nascimento (2021i), Perceptual Speed Test 2 (P2) in Gomes et al. (2021d) and Gomes and Araujo (2025a), Perceptual Speed Test 3 (P3) in Gomes et al. (2021c) and Gomes and Araujo (2025d), and Numerical Speed Test (N) in Gomes and Nascimento (2021g).

Perceptual Speed Test 1 (P1)

The respondent is asked to scan 10 columns of 41 words and mark, in each column, the five words that contain the letter "a." The test contains 50 target words and has a 2-minute time limit, so its maximum score is 50.

Perceptual Speed Test 2 (P2)

Each of the 48 items presents a pair of multi-digit numbers, and the respondent is asked to mark whether the two numbers are the same or different. The test has a 1.5-minute time limit, and its maximum score is 48.

Perceptual Speed Test 3 (P3)

Each of the 48 items presents a model figure and five options, and the respondent is asked to mark the option identical to the model. The test has a 1.5-minute time limit, and its maximum score is 48.

Numerical Speed Test (N)

The 60 items are arranged in alternating rows of 10 subtractions and 10 multiplications, each formed by two numbers with one or two digits. The test has a 2-minute time limit, and its maximum score is 60.

Procedure

The four processing speed tests were administered on paper together with the other 14 BAFACALO tests, in two 100-minute sessions that each comprised nine tests (Golino & Gomes, 2014). The Ethics Committee of the Federal University of Minas Gerais approved the data collection protocol (ETIC 181/06), and the school authorized the study. Participants aged 18 or older gave their own consent; younger participants gave their assent, and a parent or legal guardian consented on their behalf. Participants appear in the dataset only as identification numbers, so the dataset is anonymous. The dataset was published as a data

paper in the Journal of Open Psychology Data, a peer-reviewed journal that accepts only datasets deposited in public repositories under open licenses, so any researcher can cite and reuse the data (Golino & Gomes, 2014).

Data Analysis

The first step describes the distribution of the scores on each test. All analyses are conducted in R version 4.4.1 (R Core Team, 2024). Descriptive statistics (mean, standard deviation, minimum, and maximum) are computed for each test, on raw scores and on percentages, with the psych 2.4.6 package (Revelle, 2024), and figures are produced with the ggplot2 4.0.3 package (Wickham, 2016).

The scores of the four tests are counts of correct responses obtained under a time limit and are therefore analyzed with the Rasch Poisson Counts Model (RPCM; Baghaei & Doebler, 2019), with the four tests treated as indicators of a single processing speed dimension. In this model, the logarithm of the expected count of a student on a test is the sum of the easiness of the test and the processing speed of the student. Test easiness is specified as a fixed effect and processing speed as a normally distributed random effect, and the model is estimated by marginal maximum likelihood with the lme4 1.1-35.5 package (Bates et al., 2015).

The first assumption of the RPCM, that a single processing speed accounts for performance on the four tests, is assessed with two procedures. In the first, the eigenvalues of the correlation matrix of the four test scores are inspected; a dominant first eigenvalue, with the remaining eigenvalues below 1, indicates a single common dimension (Chou & Wang, 2010). In the second, a principal component analysis is applied to the Pearson residuals of the RPCM. If processing speed alone accounts for performance on the four tests, the residuals carry no common structure and no component dominates. Because of sampling error, however, a residual eigenvalue can exceed 1 even in the absence of a second dimension. The first residual eigenvalue is therefore compared with those obtained from 1,000 datasets simulated from the fitted RPCM, in which unidimensionality holds by construction. The model is refitted to each simulated dataset. Unidimensionality is supported if the observed first eigenvalue falls below the 95th percentile of the simulated eigenvalues.

The second assumption of the RPCM, that the variance of the counts equals their mean (equidispersion), is assessed with simulation-based randomized quantile residuals (Santos & Bazán, 2021), computed with the DHARMa 0.5.0 package (Hartig, 2026) from 1,000 datasets simulated from the RPCM. Each residual is the position of an observed count within the distribution of counts simulated for the same student and test, and under a correctly specified model these residuals follow a uniform distribution between 0 and 1. Equidispersion is tested with a simulation-based dispersion test, and the homogeneity of residual variance across the four tests with a Levene test. A single Kolmogorov-Smirnov (KS) test over all residuals is not used, since the KS test assumes independent and identically distributed observations (Chandy et al., 2025), and the four residuals of a student are not independent. The fit of the residuals to the uniform distribution is instead tested with a KS test for each of the four tests, in which each student contributes a single residual, at a Bonferroni-corrected significance level of .0125 (.05/4). The presence of outliers is tested with a bootstrap outlier test. In the dispersion, Levene, and outlier tests, $p < .05$ indicates

a departure from the model. If outliers are detected, the model is refitted without them to examine whether they account for the departure from equidispersion.

If equidispersion is rejected, the Poisson distribution is replaced by the Conway-Maxwell-Poisson (CMP) distribution (Forthmann et al., 2020), which keeps the Rasch structure of the RPCM and adds a dispersion parameter that allows the variance to be larger (overdispersion) or smaller (underdispersion) than the mean. Two CMP models are specified, one with a single dispersion parameter for the four tests and one with a dispersion parameter for each test. The three models are fitted with the glmmTMB 1.1.15.2 package (Brooks et al., 2017), so that their likelihoods are comparable, and are compared by likelihood ratio tests, the Akaike information criterion (AIC), and the Bayesian information criterion (BIC). In the likelihood ratio tests, each model is compared with the next more complex one, and $p < .05$ favors the more complex model. A difference of 10 or more in AIC or BIC is taken as strong evidence in favor of the model with the lower value (Burnham & Anderson, 2002; Raftery, 1995). The selected model is considered appropriate only if it passes the same assessment of the two assumptions described above.

The reliability of the processing speed measure is estimated under the final model. In Rasch models, measurement precision varies along the ability scale, so reliability is computed for each student rather than as a single coefficient (Baghaei & Doebler, 2019). The reliability for a student is the estimated variance of processing speed in the population divided by the sum of this variance and the squared standard error of the student's estimated processing speed, both on the logarithmic scale of the model and both obtained from the model fitted with glmmTMB. The distribution of these reliabilities across students is summarized by its minimum, quartiles, median, mean, and maximum.

Norms are derived from the final model, and the score on which they are based depends on which model is retained. Under the RPCM, the total score across the four tests is a sufficient statistic for processing speed, and norms are based on it. Under the CMP model with test-specific dispersion, the total score is no longer sufficient, because a test with lower dispersion carries more information about processing speed. If this model is retained, the normative score is the sum of the four test scores, each weighted by the inverse of its estimated dispersion parameter. Its adequacy is assessed by its Spearman correlation with the processing speed estimated by the model, with values close to 1 indicating that the normative score orders students as the model does. For each normative score, the mean estimated processing speed, the percentile rank, and a T score (mean 50, standard deviation 10) are computed. The conversion table is retained if the mean estimated processing speed does not decrease as the normative score increases, so that a higher normative score never corresponds to a lower T score.

RESULTS

Descriptive Statistics

The descriptive statistics are presented with a focus on what they reveal about the fit of the scores to the RPCM (Table 1). P3 was the easiest test and P2 the hardest, and the tests also differed in how much their scores varied relative to their means. The variance of P2 was the smallest relative to its mean, practically equal to it, and that of P1 the largest (SD^2/M in Table 1). In a Poisson distribution, the variance equals the mean when all students

have the same processing speed. Students differ in speed, and this difference adds variance, so under the RPCM the variance of every test exceeds its mean, and SD^2/M is greater than 1. The SD^2/M of P2 was 1.01, so its scores varied less than the RPCM predicts. In N, P1, and P3, SD^2/M exceeded 1, but how much it should exceed 1 under the RPCM depends on the spread of processing speed, which only the model estimates. The descriptive statistics therefore cannot show whether the variance of these three tests departs from the prediction. The maximum possible score was reached by 3 students (1.3%) in P1 and by 14 (6.1%) in P3. These tests cannot distinguish these students from one another, which reduces variance at the top of the distribution. In P1 this effect is negligible, whereas in P3 it is small but not negligible. The descriptive statistics thus suggest that P2, and possibly P3 among the fastest students, vary less than the RPCM predicts, which would violate its equidispersion assumption. The residual diagnostics in the next subsection assess this possibility.

Table 1: Descriptive statistics of the four speed tests (N = 230)

Test	Maximum possible score	*M*	*SD*	Min	Max	SD^2/M
N	60	52.7% (31.59)	13.5% (8.11)	21.7% (13)	96.7% (58)	2.08
P1	50	55.2% (27.60)	16.1% (8.07)	12.0% (6)	100.0% (50)	2.36
P2	48	40.3% (19.36)	9.2% (4.41)	6.2% (3)	66.7% (32)	1.00
P3	48	76.8% (36.86)	15.4% (7.40)	25.0% (12)	100.0% (48)	1.49

Note. M = mean; SD = standard deviation; Min = minimum; Max = maximum. Percentages are relative to the maximum possible score of each test, and raw scores are given in parentheses. SD^2/M = variance-to-mean ratio, computed on raw scores as the squared standard deviation (the variance) divided by the mean.

Assessment of the Model Assumptions

The RPCM rests on two assumptions, unidimensionality and equidispersion, and the four tests met the first. The first eigenvalue of the correlation matrix of the test scores was 2.07, and the remaining eigenvalues were below 1 (0.77, 0.60, and 0.56). In the residuals of the RPCM, the first eigenvalue (1.41) fell below the 95th percentile of the eigenvalues (1.45) obtained from the simulated unidimensional datasets. N, P1, P2, and P3 can therefore be treated as indicators of a single processing speed dimension. With only four tests, however, both procedures have limited power to detect a weak second dimension, so this conclusion rests on the absence of evidence against unidimensionality.

The second assumption failed: the variance of the counts departed from the equidispersion of the Poisson distribution, and the departure differed across tests. The residual variance was 1.20 times larger than expected under the RPCM ($p < .001$), and it differed across the four tests (Levene test, $F(3, 916) = 7.03$, $p < .001$). Figure 1 shows this departure in two opposite directions. Each residual is the position of an observed count within the counts the model simulates for the same student and test, so under a correct model the middle half of the residuals falls between 0.25 and 0.75 (the dashed lines). In P1, the model simulated a narrower range of counts than students actually produced, so the observed counts often fell near the extremes of the simulated counts. Their residuals moved toward 0 and 1, and the middle half widened beyond the dashed lines (from 0.22 to 0.77): the scores varied more than the RPCM predicts (overdispersion). In P2, the opposite

occurred. The model simulated a wider range of counts than students actually produced, so the observed counts rarely reached the extremes of the simulated counts and fell mostly near their center. Their residuals moved toward 0.50, and the middle half narrowed inside the dashed lines (from 0.32 to 0.71): the scores varied less than the RPCM predicts (underdispersion). Both directions violate equidispersion. The residuals of P2 also departed from the uniform distribution at the corrected level of .0125 (KS, D = .106, p = .012), whereas those of N, P1, and P3 did not (p = .91, .57, and .20, respectively). Seven observations were identified as outliers (0.76% of the observations, against 0.24% expected; p = .02), but they did not cause the departure from equidispersion, which remained after their removal (dispersion = 1.16, p < .001). The departure is thus a property of the tests and not of a few atypical students. The residuals confirmed the underdispersion of P2 anticipated by the descriptive statistics and revealed an overdispersion of P1 that the descriptive statistics could not detect.

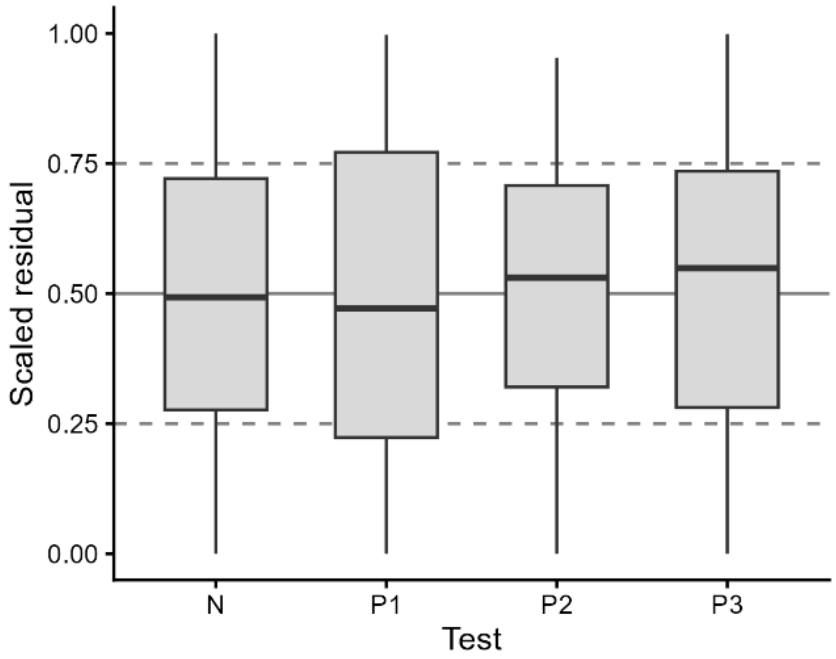


Figure 1: Distribution of the scaled residuals of each test under the Poisson model (RPCM)

Note. Under a correctly specified model, the scaled residuals follow a uniform distribution between 0 and 1. The solid line marks the expected median (0.50), and the dashed lines mark the expected quartiles (0.25 and 0.75).

Model Comparison and Final Model

Allowing each test to have its own dispersion produced the best-fitting model by all criteria (Table 2). The CMP model with a single dispersion parameter improved on the Poisson model in the likelihood ratio test, but its AIC and BIC were only 6.9 and 2.0 points lower, below the threshold of 10. The CMP model with test-specific dispersion improved on the model with a single dispersion parameter, with AIC and BIC values 37.2 and 22.8 points lower, respectively, and 44.1 and 24.8 points lower than those of the Poisson model.

Table 2: Comparison of the Poisson and Conway-Maxwell-Poisson models

Model	Parameters	Log-likelihood	AIC	BIC	x^2 (df)	p
Poisson (RPCM)	5	−3034.4	6078.9	6103.0		
CMP, single dispersion	6	−3030.0	6072.0	6101.0	8.82 (1)	.003
CMP, test-specific dispersion	9	−3008.4	6034.8	6078.2	43.22 (3)	< .001

Note. CMP = Conway-Maxwell-Poisson; AIC = Akaike information criterion; BIC = Bayesian information criterion.

Under the CMP model with test-specific dispersion, the scores met both assumptions of the RPCM. The residuals of this model preserved the evidence of a single dimension: the first residual eigenvalue (1.38) fell below the 95th percentile of the eigenvalues obtained from datasets simulated from the CMP model (1.44; median = 1.34). In P1, the middle half of the residuals narrowed from 0.22–0.77 under the Poisson model to 0.28–0.71, close to the dashed lines (Figure 2). In P2, it changed little (from 0.32–0.71 to 0.29–0.73) and remained narrower than expected, but neither the dispersion, homogeneity, nor uniformity tests detected this concentration as a departure from the model. The residual variance no longer differed from the expected variance (dispersion = 1.10, p = .036) or across tests (Levene test, F(3, 916) = 0.59, p = .62). The residuals of no test departed from the uniform distribution at the corrected level of .0125 (KS; N, p = .33; P1, p = .86; P2, p = .12; P3, p = .042), and the number of outliers did not exceed the expected number (p = .82). The CMP model with test-specific dispersion is therefore retained as the final model.

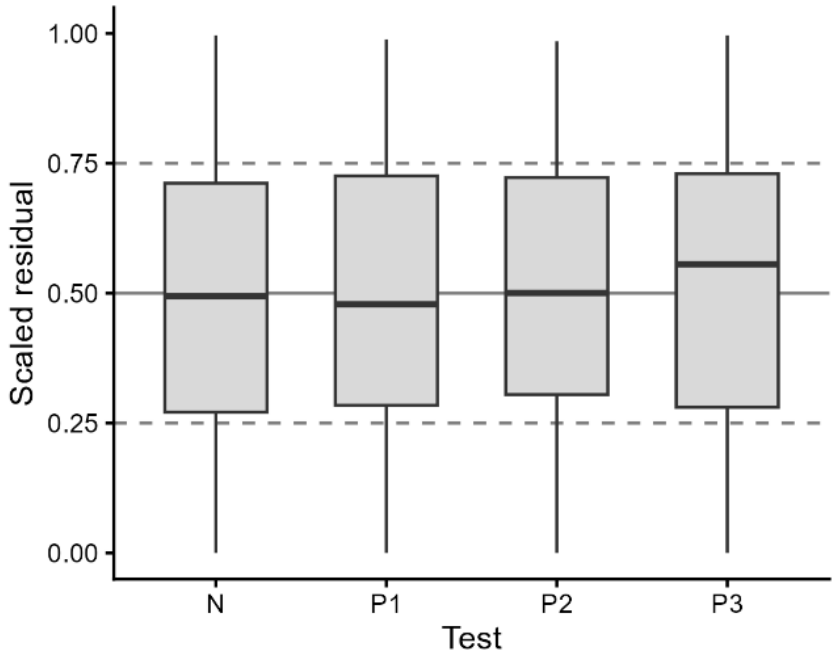


Figure 2: Distribution of the scaled residuals of each test under the CMP model with test-specific dispersion

Note. Under a correctly specified model, the scaled residuals follow a uniform distribution between 0 and 1. The solid line marks the expected median (0.50), and the dashed lines mark the expected quartiles (0.25 and 0.75).

The processing speed measure built from the final model was moderately reliable, and its reliability varied little along the speed scale. Reliability expresses how much of the variation in the estimated processing speed reflects true differences between students rather than measurement error. Across students, it ranged from .69 to .78, with a median of .75, a mean of .75, and first and third quartiles of .74 and .76, so half of the students had reliabilities within .02 of each other. The four tests thus measure students at different levels of processing speed with similar precision, so a single reliability level of about .75 describes practically the whole sample.

The Processing Speed Measure and Its Norms

In the final model, the four tests carry different amounts of information about processing speed. The dispersion parameters, where values above 1 indicate that a test varies more than the Poisson distribution allows and values below 1 that it varies less, confirmed the departures found in the residuals: P1 (1.78) and N (1.28) were overdispersed, P2 (0.64) was underdispersed, and P3 (1.03) was close to the Poisson distribution. The dispersion of a test expresses how much its scores vary for students of the same processing speed. The more a test varies for reasons other than speed, the less a correct response in it reveals about speed. A simple sum of the four scores gives every correct response the same weight and therefore no longer orders students as the model does. The normative score corrects this by dividing each test score by the dispersion of the test before summing: Normative score = N/1.28 + P1/1.78 + P2/0.64 + P3/1.03.

Each test is thus weighted by the inverse of its dispersion: 1/0.64 = 1.56 for P2, 1/1.03 = 0.97 for P3, 1/1.28 = 0.78 for N, and 1/1.78 = 0.56 for P1. A test close to the Poisson distribution, such as P3, keeps a weight close to 1, whereas the underdispersed P2 gains weight and the overdispersed P1 loses it. For example, 20 correct responses add 31.2 points to the normative score in P2 (20 × 1.56) but only 11.2 points in P1 (20 × 0.56), so a correct response in P2 counts almost three times as much as one in P1. This weighted score ordered students almost exactly as the processing speed estimated by the model (Spearman $\rho > .999$). Users of the norms can therefore locate a student by computing this score and consulting the normative table (Table 3), without fitting the model.

Table 3: Normative table for the weighted normative score (N = 230)

Score	PR	T	Score	PR	T	Score	PR	T
51.56	0.4	19.8	91.12	19.6	41.7	101.17	39.6	47.3
65.13	0.9	27.4	91.65	20.0	42.0	101.22	40.0	47.3
67.22	1.3	28.5	91.71	20.4	42.1	101.41	40.4	47.4
69.08	1.7	29.5	91.81	20.9	42.1	101.57	40.9	47.5
71.32	2.2	30.8	91.96	21.3	42.2	101.72	41.3	47.6
71.48	2.6	30.9	92.38	21.7	42.4	101.73	41.7	47.6
71.50	3.0	30.9	92.76	22.2	42.6	101.99	42.2	47.8
74.09	3.5	32.3	93.12	22.6	42.8	102.18	42.6	47.9
74.16	3.9	32.4	93.23	23.0	42.9	102.44	43.0	48.0
75.30	4.3	33.0	93.82	23.9	43.2	102.53	43.5	48.1
75.71	4.8	33.2	94.30	24.3	43.5	102.68	43.9	48.1
76.20	5.2	33.5	94.48	24.8	43.6	102.90	44.3	48.3

76.58	5.7	33.7	94.51	25.2	43.6	103.55	44.8	48.6
76.93	6.1	33.9	94.92	25.7	43.8	103.73	45.2	48.7
77.20	6.5	34.0	94.96	26.1	43.9	103.80	45.7	48.8
77.71	7.0	34.3	95.17	26.5	44.0	103.84	46.1	48.8
78.12	7.4	34.5	95.50	27.0	44.2	104.01	46.5	48.9
78.32	7.8	34.7	95.53	27.8	44.2	104.03	47.0	48.9
78.70	8.3	34.9	95.73	28.3	44.3	104.21	47.4	49.0
79.25	8.7	35.2	95.88	28.7	44.4	104.25	47.8	49.0
79.85	9.1	35.5	96.16	29.1	44.5	104.44	48.3	49.1
80.28	9.6	35.7	96.40	29.6	44.7	104.77	48.7	49.3
80.85	10.0	36.1	96.54	30.0	44.7	105.24	49.1	49.6
82.00	10.4	36.7	97.07	30.4	45.0	105.34	49.6	49.6
82.16	10.9	36.8	97.39	30.9	45.2	105.51	50.0	49.7
83.01	11.3	37.3	97.42	31.3	45.2	105.57	50.4	49.7
83.78	11.7	37.7	97.50	31.7	45.3	105.59	50.9	49.7
85.13	12.2	38.4	97.59	32.2	45.3	106.08	51.3	50.0
85.38	12.6	38.6	98.03	32.6	45.6	106.11	51.7	50.0
86.10	13.0	39.0	98.30	33.0	45.7	106.29	52.2	50.1
86.36	13.5	39.1	98.41	33.5	45.8	106.51	52.6	50.3
86.48	13.9	39.2	98.59	33.9	45.9	106.75	53.0	50.4
86.79	14.3	39.3	98.68	34.3	45.9	107.14	53.5	50.6
87.68	14.8	39.8	98.70	34.8	45.9	107.32	53.9	50.7
87.76	15.2	39.9	98.78	35.2	46.0	107.38	54.3	50.7
87.85	15.7	39.9	99.22	35.7	46.2	108.02	54.8	51.1
87.93	16.1	40.0	99.98	36.1	46.6	108.08	55.2	51.1
88.50	16.5	40.3	100.29	36.5	46.8	108.17	55.7	51.2
88.60	17.0	40.3	100.30	37.0	46.8	108.28	56.1	51.2
88.84	17.4	40.5	100.34	37.4	46.8	108.35	56.5	51.3
90.26	17.8	41.3	100.35	37.8	46.8	108.38	57.0	51.3
90.56	18.3	41.4	100.38	38.3	46.9	109.18	57.4	51.7
90.68	18.7	41.5	100.81	38.7	47.1	109.29	57.8	51.8
90.78	19.1	41.6	101.11	39.1	47.3	109.30	58.3	51.8
109.63	58.7	52.0	117.63	72.6	56.4	126.65	86.5	61.4
109.67	59.1	52.0	117.83	73.0	56.5	127.23	87.0	61.7
110.05	59.6	52.2	117.91	73.5	56.6	127.28	87.4	61.8
110.06	60.0	52.2	118.10	73.9	56.7	127.43	87.8	61.8
110.16	60.4	52.3	118.29	74.3	56.8	127.53	88.3	61.9
110.21	60.9	52.3	118.40	74.8	56.8	127.69	88.7	62.0
110.56	61.3	52.5	118.65	75.2	57.0	128.33	89.1	62.3
110.60	61.7	52.5	118.89	75.7	57.1	128.34	89.6	62.3
110.75	62.2	52.6	118.98	76.1	57.2	128.87	90.0	62.6
110.88	62.6	52.7	119.07	76.5	57.2	128.91	90.4	62.7
111.02	63.0	52.8	119.50	77.0	57.4	129.55	90.9	63.0
111.35	63.5	52.9	119.51	77.4	57.5	129.91	91.3	63.2
111.56	63.9	53.1	119.93	77.8	57.7	130.17	91.7	63.4
111.76	64.3	53.2	121.29	78.3	58.4	130.22	92.2	63.4
111.80	64.8	53.2	121.38	78.7	58.5	130.73	92.6	63.7
111.88	65.2	53.2	121.70	79.1	58.7	130.88	93.0	63.7
112.00	65.7	53.3	121.75	79.6	58.7	131.14	93.5	63.9
112.04	66.1	53.3	121.88	80.0	58.8	131.56	93.9	64.1

112.08	66.5	53.3	122.43	80.4	59.1	134.78	94.3	65.9
112.68	67.0	53.7	122.79	80.9	59.3	135.46	94.8	66.3
113.46	67.4	54.1	123.03	81.3	59.4	135.56	95.2	66.3
113.69	67.8	54.2	123.29	81.7	59.5	136.09	95.7	66.6
113.91	68.3	54.4	123.62	82.2	59.7	136.21	96.1	66.7
114.41	68.7	54.6	123.88	82.6	59.9	137.63	96.5	67.5
114.89	69.1	54.9	124.37	83.0	60.1	137.77	97.0	67.6
115.05	69.6	55.0	124.76	83.5	60.4	139.19	97.4	68.3
115.40	70.0	55.2	124.80	83.9	60.4	139.32	97.8	68.4
115.68	70.4	55.3	125.29	84.3	60.7	140.01	98.3	68.8
115.97	70.9	55.5	125.53	84.8	60.8	148.24	98.7	73.4
116.23	71.3	55.6	126.20	85.2	61.2	151.62	99.1	75.2
116.34	71.7	55.7	126.57	85.7	61.4	154.59	99.6	76.9
117.18	72.2	56.2	126.63	86.1	61.4	156.98	100.0	78.2

Note. Score = normative score, the sum of the four raw test scores, each divided by its estimated dispersion parameter (N, 1.278; P1, 1.784; P2, 0.642; P3, 1.032). PR = percentile rank, the percentage of students with a normative score equal to or lower than the given score. T = T score, the normative score standardized to a mean of 50 and a standard deviation of 10. For a score not listed, use the nearest lower score.

DISCUSSION

The four tests depart from the Poisson distribution in different directions because of how their scored responses accumulate over time. A Poisson count arises when the waiting time between successive responses is as irregular as an exponential distribution implies. When the waiting times are more regular, the count varies less than the Poisson distribution predicts, and when they are more irregular, it varies more (Winkelmann, 1995). In P2, every item requires the same operation, a single same-or-different judgment about two strings of digits, so students work at a steady pace, which is consistent with the underdispersion of P2 (0.64). In P1, only the words containing the letter A are scored, but students must read every word, and the five targets of each column of 41 words are irregularly spaced among the nontargets. The time between two scored responses therefore depends on how many nontargets separate them, and an overlooked target lowers the score without being detected, so P1 varies for reasons other than processing speed, which is consistent with its overdispersion (1.78). The items of N progress from one-digit subtractions to multiplications of two-digit by one-digit numbers, so performance also depends on arithmetic fluency, a source of individual differences specific to this test that processing speed does not capture, which is consistent with the overdispersion of N (1.28). P3, a figure-matching task in which every item requires a similar visual comparison among five alternatives, was close to the Poisson distribution (1.03). These explanations are hypotheses derived from the structure of the tests. The present data do not test them, and response times for each item would be needed to do so.

Dispersion that differs in direction across the parts of a measure has been reported for other count-based cognitive tasks, and the present results extend this pattern to perceptual and numerical speed tests. In verbal fluency data, a CMP model with item-specific dispersion fitted best, with three of four items underdispersed (Forthmann et al., 2020), and a divergent thinking measure contained both overdispersed and underdispersed items (Beisemann, 2022). In these studies, as in the present one, the Poisson assumption

failed not only in degree but also in direction, so a single dispersion parameter for all items or tests could not represent the data.

Testing equidispersion is therefore not a formality in the RPCM, since its outcome changes the score assigned to each student. Under the Poisson model, the total score is sufficient, and a correct response counts the same in every test. Had the analysis stopped at the RPCM, the norms would rest on this simple sum, although a correct response in P2 carries almost three times the information of a correct response in P1. Departures from equidispersion also bias the reliability estimated under the RPCM (Forthmann et al., 2020). Studies that apply the RPCM to speed tests should therefore assess dispersion for each test and, when the tests differ, base the scores on a model that represents these differences.

Some limitations qualify these conclusions. The weights of the normative score are dispersion estimates from 230 students and may change in other samples, and the normative score is an empirical approximation to the model rather than a sufficient statistic derived from it. The same sample size limits the precision of the percentiles at the extremes of the distribution. Replicating the dispersion pattern and the weights in an independent sample is the next step before the norms are used in practice.

CONCLUSION

This study contributes to psychometrics by applying a Rasch count model. These models are rarely used, but they suit speed tests well, because in such tests the items are very easy and the time limit is what constrains performance. Modeling the counts of correct responses made it possible to build a measure of the broad ability of processing speed (Gs) from the BAFACALO tests P1, P2, P3, and N.

The RPCM was the first model tested because, under this model, the scores on the four tests could be summed and the sum applied directly to the norms of the measure. However, the scores violated one of its assumptions, equidispersion. Even so, the CMP model with test-specific dispersion, selected in this study, also yielded norms that are easy to use. Under this model, the scores on the four tests can still be summed and the sum applied to the norms, provided that each test score is first weighted by the simple calculation presented in this study.

The evidence in this study, however, comes from a small sample of students from a single school. New studies with larger and more diverse samples are therefore needed.

DECLARATIONS

Funding: None.

Conflicts of Interest: None.

Data Availability: The data are openly available in Golino and Gomes (2014; https://doi.org/10.5334/jopd.af).

Author Contributions (CRediT): Both authors contributed to all stages of the study.

Ethics & Consent: The original data collection was approved by the UFMG Ethics Committee (ETIC 181/06), with informed consent.

Use of Generative AI: Used for writing support, mainly English-language editing.

Acknowledgements: Cristiano Mauro Assis Gomes holds a Research Productivity Fellowship from the National Council for Scientific and Technological Development (CNPq), Brazil.

REFERENCES

Alves, F. A., Flores, R. P., Gomes, C. M. A., & Golino, H. F. (2012). Preditores do rendimento escolar: Inteligência geral e crenças sobre ensino-aprendizagem. Revista E-Psi, 1, 97–117. https://www.researchgate.net/publication/235642286_Preditores_do_Rendimento_Escolar_Inteligencia_Geral_e_Crencas_sobre_Ensino-Aprendizagem

Baghaei, P., & Doebler, P. (2019). Introduction to the Rasch Poisson counts model: An R tutorial. Psychological Reports, 122(5), 1967–1994. https://doi.org/10.1177/0033294118797577

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beisemann, M. (2022). A flexible approach to modelling over-, under- and equidispersed count data in IRT: The Two-Parameter Conway–Maxwell–Poisson Model. British Journal of Mathematical and Statistical Psychology, 75(3), 411–443. https://doi.org/10.1111/bmsp.12273

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. The R Journal, 9(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Burnham, K. P., & Anderson, D. R. (2002). Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.). Springer. https://doi.org/10.1007/b97636

Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytic studies. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Chandy, M., Schifano, E. D., Yan, J., & Zhang, X. (2025). Nonparametric block bootstrap Kolmogorov-Smirnov goodness-of-fit test (arXiv:2511.05733). arXiv. https://doi.org/10.48550/arXiv.2511.05733

Chou, Y.-T., & Wang, W.-C. (2010). Checking dimensionality in item response models with principal component analysis on standardized residuals. Educational and Psychological Measurement, 70(5), 717–731. https://doi.org/10.1177/0013164410379322

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., Orben, A., Parsons, S., & Schulte-Mecklenbeck, M. (2019). Seven easy steps to open science: An annotated reading list. Zeitschrift für Psychologie, 227(4), 237–248. https://doi.org/10.1027/2151-2604/a000387

Forthmann, B., Gühne, D., & Doebler, P. (2020). Revisiting dispersion in count data item response theory models: The Conway–Maxwell–Poisson counts model. British Journal of Mathematical and Statistical Psychology, 73(S1), 32–50. https://doi.org/10.1111/bmsp.12184

Golino, H. F., & Gomes, C. M. A. (2014). Psychology data from the "BAFACALO project: The Brazilian Intelligence Battery based on two state-of-the-art models – Carroll's Model and the CHC model". Journal of Open Psychology Data, 2(1), e6. https://doi.org/10.5334/jopd.af

Gomes, C. M. A. (2010a). Avaliando a avaliação escolar: Notas escolares e inteligência fluida. Psicologia em Estudo, 15(4), 841–849. https://doi.org/10.1590/S1413-73722010000400020

Gomes, C. M. A. (2010b). Estrutura fatorial da Bateria de Fatores Cognitivos de Alta-Ordem (BaFaCalo). Avaliação Psicológica, 9(3), 449–459. http://pepsic.bvsalud.org/scielo.php?pid=S1677-04712010000300011&script=sci_arttext

Gomes, C. M. A. (2011). Validade do conjunto de testes da habilidade de memória de curto-prazo (CTMC). Estudos de Psicologia (Natal), 16(3), 235–242. https://doi.org/10.1590/S1413-294X2011000300005

Gomes, C. M. A. (2012). Validade de construto do conjunto de testes de inteligência cristalizada (CTIC) da bateria de fatores cognitivos de alta-ordem (BaFaCAlO). Gerais: Revista Interinstitucional de Psicologia, 5(2), 294–316. http://pepsic.bvsalud.org/scielo.php?pid=S1983-82202012000200009&script=sci_arttext

Gomes, C. M. A., & Araujo, J. (2025a). Ciência aberta e democratização de testes: Disponibilização gratuita e aberta do Teste de Velocidade Perceptiva 2 da Bateria de Fatores Cognitivos de Alta Ordem (BAFACALO). Revista Contemporânea, 5(2), e7431. https://doi.org/10.56083/RCV5N2-038

Gomes, C. M. A., & Araujo, J. (2025b). Democratizando os testes de inteligência: Acesso aberto e gratuito ao Teste de Dobraduras da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO). European Journal of Education Studies, 12(1), 109–136. https://doi.org/10.46827/ejes.v12i1.5771

Gomes, C. M. A., & Araujo, J. (2025c). Democratizando os testes de inteligência: Acesso aberto e gratuito ao Teste de Indução da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO). Observatorio de la Economía Latinoamericana, 23(1), 1–28. https://doi.org/10.55905/oelv23n1-136

Gomes, C. M. A., & Araujo, J. (2025d). Teste de Velocidade Perceptiva 3 da Bateria de Fatores Cognitivos de Alta Ordem (BAFACALO): Disponibilização gratuita e aberta. Cuadernos de Educación y Desarrollo, 17(3), e7727. https://doi.org/10.55905/cuadv17n3-031

Gomes, C. M. A., & Araujo, J. (2026a). Incorporating test design into factorial validity: A bifactor reanalysis of the BAFACALO Folding Test. REMUNOM, 13(21), 1–17. https://doi.org/10.66104/00fx1964

Gomes, C. M. A., & Araujo, J. (2026b). Modeling Response Processes to Achieve an Appropriate Factor Structure for the Bafacalo Logical Reasoning Test. Advances in Social Sciences Research Journal, 13(10), 366-378. https://doi.org/10.14738/assrj.1310.12492

Gomes, C. M. A., & Borges, O. N. (2009a). Propriedades psicométricas do conjunto de testes da habilidade visuoespacial. Psico-USF, 14(1), 19–34. https://doi.org/10.1590/S1413-82712009000100004

Gomes, C. M. A., & Borges, O. (2009b). Qualidades psicométricas do conjunto de testes de inteligência fluida. Avaliação Psicológica, 8(1), 17–32. http://pepsic.bvsalud.org/scielo.php?pid=S1677-04712009000100003&script=sci_arttext

Gomes, C. M. A., de Araújo, J., Ferreira, M. G., & Golino, H. F. (2014). The validity of the Cattell-Horn-Carroll model on the intraindividual approach. Behavioral Development Bulletin, 19(4), 22–30. https://doi.org/10.1037/h0101078

Gomes, C. M. A., Farias, H. B., & Jelihovschi, E. G. (2022). Approaches to learning does matter to predict academic achievement. Revista de Psicología, 40(2), 905–933. https://doi.org/10.18800/psico.202202.010

Gomes, C. M. A., & Golino, H. F. (2012a). O que a inteligência prediz: Diferenças individuais ou diferenças no desenvolvimento acadêmico? Psicologia: Teoria e Prática, 14(1), 126–139. http://pepsic.bvsalud.org/pdf/ptp/v14n1/v14n1a10.pdf

Gomes, C. M. A., & Golino, H. F. (2012b). Validade incremental da Escala de Abordagens de Aprendizagem (EABAP). Psicologia: Reflexão e Crítica, 25(4), 400–410. https://doi.org/10.1590/S0102-79722012000400001

Gomes, C. M. A., & Golino, H. (2015). Factor retention in the intra-individual approach: Proposition of a triangulation strategy. Avaliação Psicológica, 14(2), 273–279. https://doi.org/10.15689/ap.2015.1402.12

Gomes, C. M. A., Marques, E. L. L., & Golino, H. F. (2014). Validade incremental dos estilos legislativo, executivo e judiciário em relação ao rendimento escolar. Revista E-Psi, 2, 31–46. https://revistaepsi.com/artigo/2013-2014-ano3-volume2-artigo3/

Gomes, C. M. A., & Nascimento, D. F. (2021a). A medida da habilidade de fluência do modelo CHC: Apresentando o Teste de Fluência Ideativa 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.35726.28481/1

Gomes, C. M. A., & Nascimento, D. F. (2021b). Acesso aberto ao Teste de Fluência Figural da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) como medida da habilidade ampla de fluência do modelo CHC de inteligência [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.15593.62564/1

Gomes, C. M. A., & Nascimento, D. F. (2021c). Acesso aberto e gratuito ao Conjunto de Testes de Inteligência Fluida: Teste de Raciocínio Geral da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.30509.61921/1

Gomes, C. M. A., & Nascimento, D. F. (2021d). Acesso aberto e gratuito ao Teste de Fluência Ideativa 1 da BAFACALO [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.24821.09442/3

Gomes, C. M. A., & Nascimento, D. F. (2021e). Apresentando o Teste de Flexibilidade de Fechamento da BAFACALO [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.31920.28164

Gomes, C. M. A., & Nascimento, D. F. (2021f). Disponibilizando de forma gratuita e aberta o Teste de Memória Associativa 1 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.29964.03201/1

Gomes, C. M. A., & Nascimento, D. F. (2021g). Disponibilizando de forma gratuita e aberta o Teste de Velocidade Numérica da BAFACALO [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.24114.94407/1

Gomes, C. M. A., & Nascimento, D. F. (2021h). Medidas de inteligência cristalizada: Disponibilizando o Teste de Compreensão Verbal 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.36085.09447/1

Gomes, C. M. A., & Nascimento, D. F. (2021i). Medindo a habilidade de rapidez cognitiva do modelo CHC: Apresentando o Teste de Velocidade Perceptiva 1 da BAFACALO [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.28564.83848/1

Gomes, C. M. A., & Nascimento, D. F. (2021j). Projeto de acesso aberto e gratuito à Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO): O Teste de Compreensão Verbal 1 do Conjunto de Testes de Inteligência Cristalizada [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.22663.32165/1

Gomes, C. M. A., & Nascimento, D. F. (2021k). Projeto de acesso aberto e gratuito aos testes do LAICO: Teste de Raciocínio Lógico da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.25476.45445/1

Gomes, C. M. A., & Nascimento, D. F. (2021l). Projeto de acesso aos testes de inteligência da BAFACALO: Teste de Compreensão Verbal 3 [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.10499.84001/2

Gomes, C. M. A., & Nascimento, D. F. (2021m). Projeto de acesso da BAFACALO: Teste de Memória Associativa 2 [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.23253.14565/1

Gomes, C. M. A., & Nascimento, D. F. (2021n). Teste de Memória Visual da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.33319.47529

Gomes, C. M. A., Nascimento, D. F., & Araujo, J. (2021a). Acesso aberto ao Teste de Dobraduras (VZ) da BAFACALO [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.21853.95201/2

Gomes, C. M. A., Nascimento, D. F., & Araujo, J. (2021b). Medindo a inteligência fluida: O Teste de Indução da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO) [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.17087.84641/3

Gomes, C. M. A., Nascimento, D. F., & Araujo, J. (2021c). Projeto de testes gratuitos e abertos do LAICO: Teste de Velocidade Perceptiva 3 da BAFACALO [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.36278.42563/2

Gomes, C. M. A., Nascimento, D. F., & Araujo, J. (2021d). Teste de Velocidade Perceptiva 2 da Bateria de Fatores Cognitivos de Alta-Ordem (BAFACALO): Disponibilização aberta e gratuita aos testes de medida de rapidez cognitiva do LAICO [Preprint]. ResearchGate. https://doi.org/10.13140/RG.2.2.29567.53928/1

Hartig, F. (2026). DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models (Version 0.5.0) [R package]. https://CRAN.R-project.org/package=DHARMa

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Promoting an open research culture. Science, 348(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

R Core Team. (2024). R: A language and environment for statistical computing (Version 4.4.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Raftery, A. E. (1995). Bayesian model selection in social research. Sociological Methodology, 25, 111–163. https://doi.org/10.2307/271063

Revelle, W. (2024). psych: Procedures for psychological, psychometric, and personality research (Version 2.4.6) [R package]. Northwestern University. https://CRAN.R-project.org/package=psych

Santos, N. C. A., & Bazán, J. L. (2021). Residual analysis in Rasch counts models. In M. Wiberg, D. Molenaar, J. González, U. Böckenholt, & J.-S. Kim (Eds.), Quantitative psychology: The 85th Annual Meeting of the Psychometric Society, Virtual, 2020 (pp. 285–295). Springer. https://doi.org/10.1007/978-3-030-74772-5_26

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (4th ed., pp. 73–163). The Guilford Press.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-24277-4

Winkelmann, R. (1995). Duration dependence and dispersion in count-data models. Journal of Business & Economic Statistics, 13(4), 467–474. https://doi.org/10.1080/07350015.1995.10524620