Neuropsychiatric Disease and Treatment

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Latent class analysis of reading, decoding, and writing performance using the Academic Performance Test: concurrent and discriminating validity

This article was published in the following Dove Press journal:
Neuropsychiatric Disease and Treatment
13 August 2013
Number of times this article has been viewed

Hugo Cogo-Moreira[1]
Carolina Alves Ferreira Carvalho[2]
Adriana de Souza Batista Kida[2]
Clara Regina Brandão de Avila[2]
Giovanni Abrahão Salum[3,5]
Tais Silveira Moriyama[1,4]
Ary Gadelha[1,5]
Luis Augusto Rohde[3,5]
Luciana Monteiro de Moura[1]
Andrea Parolin Jackowski[1]
Jair de Jesus Mari[1]

[1]Department of Psychiatry, Federal University of São Paulo, São Paulo, [2]Department of Hearing and Speech Pathology, Federal University of São Paulo, São Paulo, [3]Department of Psychiatry, Federal University of Rio Grande do Sul, Rio Grande do Sul, [4]Department of Psychiatry, University of São Paulo, São Paulo, [5]National Institute for Developmental Psychiatry for Children and Adolescent, (National Counsel of Technological and Scientific Development), Brazil

Correspondence: Hugo Cogo-Moreira
Department of Psychiatry, Federal University of São Paulo, Rua Borges Lagoa, 570 1°Andar, São Paulo, Brazil
Tel +55 11 982 083 526
Email hugocogobr@gmail.com

Aim: To explore and validate the best returned latent class solution for reading and writing subtests from the Academic Performance Test (TDE).

Sample: A total of 1,945 children (6–14 years of age), who answered the TDE, the Development and Well-Being Assessment (DAWBA), and had an estimated intelligence quotient (IQ) higher than 70, came from public schools in São Paulo (35 schools) and Porto Alegre (22 schools) that participated in the 'High Risk Cohort Study for Childhood Psychiatric Disorders' project. They were on average 9.52 years old (standard deviation = 1.856), from the 1st to 9th grades, and 53.3% male. The mean estimated IQ was 102.70 (standard deviation = 16.44).

Methods: Via Item Response Theory (IRT), the highest discriminating items ('a'>1.7) were selected from the TDE subtests of reading and writing. A latent class analysis was run based on these subtests. The statistically and empirically best latent class solutions were validated through concurrent (IQ and combined attention deficit hyperactivity disorder [ADHD] diagnoses) and discriminant (major depression diagnoses) measures.

Results: A three-class solution was found to be the best model solution, revealing classes of children with good, not-so-good, or poor performance on TDE reading and writing tasks. The three-class solution has been shown to be correlated with estimated IQ and to ADHD diagnosis. No association was observed between the latent class and major depression.

Conclusion: The three-class solution showed both concurrent and discriminant validity. This work provides initial evidence of validity for an empirically derived categorical classification of reading, decoding, and writing performance using the TDE. A valid classification encourages further research investing correlates of reading and writing performance using the TDE.

Keywords: Academic Performance Test, TDE, decoding, writing, validity

Introduction

The assessment of word recognition and spelling is considered to be, in different alphabetic languages, an effective approach to estimating the performance of basic cognitive levels of processing related to decoding and encoding the writing system. The scores generated by the use of quantitative parameters of errors, both in reading and writing tests, allow inferences about the characteristics of cognitive processes linked to the principle of the alphabetic-orthographic system.[1] New reliable and valid data in this area of study are crucial for improving our knowledge in this field and for helping clinicians, researchers, and policymakers; diagnoses of learning disorders and language difficulties in low- and middle-income countries depend on extremely scarce scientific data to drive an evidence-based framework.

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/NDT.S45785

The Academic Performance Test (TDE)[2] is a standardized and validated test for the Brazilian population that assesses both the performance of reading aloud and writing by dictation which are known as the subtests Reading and Writing, respectively. These subtests seek to eliminate interference as much as possible in interpretation and understanding using isolated items (single words). The TDE was created approximately 10 years ago and has been frequently used to measure the performance of basic skills learned in school. It is also one of the instruments used in the diagnosis of learning disorders.

Despite its frequent use, the literature on the test so far is limited in a number of important ways. First, there are very few studies investigating the validity of the TDE among the Brazilian population, and the studies available rely on small samples and lack an assessment of concurrent mental disorders and a proper characterization of cognition. Second, recent research has pointed out some concerns related to discrimination, item difficulty, and the original criteria of the test.[3] To our knowledge, no study has investigated the properties of the TDE using modern statistical techniques such as Item Response Theory (IRT). Third, as far as we are aware, there is no study investigating the TDE from a person-centered perspective (eg, Latent Class Analysis [LCA]), which allows us to investigate classes of subjects that are distinct from each other in both writing and decoding performance simultaneously.

LCA is a form of cluster analysis initially introduced by Lazarsfeld and Henry in 1968.[4] It is the most commonly applied latent structure model for categorical data,[5] allowing the specification of statistical distributions through a model-based method, which differs from methods that apply arbitrary distance metrics to group individuals based on their similarity (for example, K-means clustering).[6] In LCA, unlike in K-means clustering, a statistical model is built for the population from which the data sample was obtained.[7] Latent class analysis may be used as a way to evaluate the diagnostic accuracy of a test when there is not a gold standard against which to compare it.[8] The latent class refers to grouping by performance standards in all the items that measure them. These groupings should then be validated to confirm whether the instrument assesses what it proposes. Also, for a proper analysis of basic skills in the domain of the alphabetic principle, it is important to consider performance in both writing and reading simultaneously.

Here, we make use of a large sample of children to investigate the suitability of test items and to explore and validate classes of subjects with problems in reading and writing simultaneously. It was hypothesized that if there is an underlying latent grouping of children, without a priori classification of the TDE's decoding (word-level skill task) and writing (dictation) items, such categories will have concurrent and divergent validity with other gold standard measures. In other words, if the best latent class solution is a useful screening tool for assessing decoding and writing skills (via very high discrimination of the TDE's items), latent groups of readers showing different levels of decoding and writing skills will be found. Indeed, if the children, who were classified as having different levels of reading and writing skills, show corresponding performance in direct measures of reading, this result will be taken as concurrent validity for the instrument. Conversely, discriminant validity will be granted if no associations are found between reading and writing ability and measurements that are not directly related to reading and writing skills.

Materials and methods

This report is part of a large, community school-based study, performed in multiple steps combining standardized evaluation from a psychiatric and cognitive neuroscience perspective, genetics, and neuroimaging to inform preventive strategies in developmental psychiatry. Our sample in the screening phase consisted of students in public schools with more than 10,000 students in the age range assessed from 57 public schools, located close to the research centers in Porto Alegre and São Paulo, Brazil. The methodological aspects of the project were extensively reviewed by Salum et al;[9] they are briefly presented here: (1) screening; (2) psychiatric assessments; and (3) cognitive evaluation, where assessments of IQ and decoding and writing skills (via the TDE) were collected. This study was approved by the ethics committee of the University of São Paulo (IORG0004884, project Institutional Review Board registration number: 1132/08). Written consent was obtained from all the parents of the participants, and verbal assent was obtained from all the children. When appropriate, written assent also was obtained from the children.

Participants

Our sample was consisted of 1,945 children, who were evaluated by hearing and speech therapists, psychologists, and psychiatrists. They came from São Paulo (35 schools) and Porto Alegre (22 schools) which participated in screening and enrollment procedures; students were on average 9.52 years old (standard deviation [SD] = 1.856), in 1st grade to 9th grade, 53.30% male, and had an intelligence quotient (IQ) higher than 70 (the sample's average IQ was 102.70 [SD = 16.44]). Only children with an estimated IQ higher than 70 were considered in order to avoid intellectual disabilities,

which might be a potential confounder for low achievement in decoding and writing (word-level skills); the proposed cutoff for intellectual disability (<70) was based on both of the most commonly used diagnostics systems in psychiatry, the DSM-IV (Diagnostic Manual of Mental Disorders) and ICD-10 (International Classification of Diseases).

Of the 1,945 children, 772 children were randomly selected from the population, and 1,173 came from the high-risk strata. Selection for the high-risk group involved a risk-prioritization procedure that focused on individuals with a family history of a disorder and/or ongoing symptoms in one of the five targeted domains (attention deficit hyperactivity disorder [ADHD], anxiety, obsessive compulsive disorder [OCD], psychosis, major depression, and learning disorders), as detected during screening.[9]

Academic Performance Test

The Academic Performance Test (TDE)[2] is composed of three subtests: writing (isolated words in dictation); mathematics (oral problem solving and written calculations of mathematical operations); and reading (recognition of isolated words). The instrument is constructed and validated for the Brazilian population.[2] In this research, only the writing and reading subtests were used. The TDE was administered by trained hearing and speech therapists, who gave the following instructions to the children: (a) for the reading subtest: "look at these words carefully and read them aloud"; and (b) for the dictation subtest: "now we are going to do a dictation. I will dictate a word. After that, I will read a sentence where this word appears, and, thus, I will read it [the word] once again. If there is some word which you do not know, try to write it in a way that you know."

Training procedures

Training procedures consisted of two sessions with experienced professionals in the field. Full explanations about the project, instruments, procedures, and standardization for the clinical evaluations were provided, in order to avoid any assessment bias.

Intelligence

IQ was estimated using the vocabulary and block design subtests of the Weschler Intelligence Scale for Children, 3rd edition (WISC-III),[10] the Tellegen and Briggs method which is a modified part-whole correlation formula offered for use in the case of non-independent test administration of the part-Wechsler subtest combination,[11] and Brazilian normalizations.[12] The WISC-III was administered by trained psychologists.

Potential confounder

Simplified Auditory Assessments[13] were conducted by a hearing and speech pathologist. These assessments tested the elicitation of the auropalpebral reflex through instrumental sounds; sound location in five directions; sequential verbal memory for sounds with three and four syllables; and sequential nonverbal memory with three and four percussion musical instruments. The children were classified as having problems in auditory processing, or not. Auditory perception may be one source of individual variation in the phonological abilities that play a critical role in skilled reading, as well as in reading and writing disabilities.[14]

Also, in order to evaluate the children's socioeconomic status (SES), the Brazilian Association of Research Companies (ABEP)[15] questionnaire was used, which measures personal material wealth and is a standardized index of economic classification, based on the family's power of consumption.

Psychiatric diagnosis

Child psychiatric diagnosis was established using the Development and Well-Being Assessment (DAWBA),[16] which was administered to the biological parents of all children included in the project. The DAWBA is a structured interview administered by lay interviewers who also record verbatim responses, which were used to confirm or refute diagnoses of any reported problems. All questions are closely related to DSM-IV diagnostic criteria and focus on current problems causing significant distress or social impairment.

A total of nine well-trained psychiatrists performed the rating procedures; they were trained and supervised closely by a senior child psychiatrist with extensive experience in rating the DAWBA. Cases about which raters had doubts about any specific diagnosis were scaled up and discussed between two child psychiatrists until consensus about the diagnosis was achieved.

Statistical analysis plan

First, in order to identify the highest discriminating TDE items, IRT was conducted (we began with this procedure, because the TDE has showed floor and ceiling effect for some items). Through this technique, it is possible to estimate the amount of latent trait (ability) required to correctly answer the item (here, individually, two different latent traits were considered: decoding [word-level reading] and writing [word-level writing], following IRT's unidimensionality assumption). All 104 items are dichotomous (ie, had only a correct or wrong answer). In order to satisfy

the unidimensionality's postulate, we admitted that there is a dominant ability (latent trait) that is measurable by achievement in a conjunction of items in each of the TDE's subtests.[17]

The greater the ability level, the higher the probability of a correct response, and vice versa. Indeed, the probability of success depends on three items' parameters: discrimination (denoted by 'a', which describes how well an item can differentiate between examinees having abilities below the item location and those having abilities above the item location);[18] difficulty (denoted by 'b', where typical values have the range $-3 \leq b \leq 3$, designating the difficulty of the item; in other words, it describes where the item functions along the ability scale); and guessing (normally used in multiple-choice tests; in other words, it is the probability of getting the item correct by guessing alone. In this research, the guessing parameter was not considered, because we are assuming that its probability is 0). Therefore, we worked using a two-parameter logistic model. Baker has pointed out that items with very high discrimination value have 'a'>1.7;[18] based on this criterion, in each latent trait, we looked for items with very high values on the discrimination parameter, which means that the item has a high power of differentiation (discrimination) between children with and without word-level reading and writing skills.

Subsequently, the set of high discriminative items were analyzed via latent class analysis (LCA), a mixture model that aims to uncover unobserved heterogeneity in a population and to find substantively meaningful groups of people that are similar in their responses to measured variables.[19] The idea behind the latent class was to find latent groups underlying the highest discriminative items (ie, items with 'a'>1.7); in other words, latent groups under items that differentiate between children having abilities below the item location and those having abilities above the item location.

To compare different numbers of latent classes' solutions, different information criterion indices were used: Akaike's Information Criterion (AIC); the Bayesian information criterion (BIC), in which small values correspond to better fit; and the sample size-adjusted BIC (ssaBIC). The Lo–Mendell–Rubin (LMR) test[20,21] and bootstrapped likelihood ratio test (BLRT) were used to test the number of classes in this mixture analysis procedure; the former is obtained by running the k-class and k-1 class analyses and using the derivatives from both models to compute the *P*-value (a low *P*-value rejects the k-1 class model in favor of the k-class model). The latter is obtained by bootstrapping, following the procedure described by Asparouhov and Muthén.[22] The classification quality of the model was evaluated according to the entropy criterion, in which the values range from 0 to 1, where values close to 1 indicate good classification. IRT and LCA were conducted via Mplus (v6.12; Muthén and Muthén, Los Angeles, CA, USA).[23]

In order to validate the best latent class solution (based on statistical and empirical evidence), concurrent and discriminant validity were assessed using the regression model from STATA (v12; StataCorp, College Station, TX, USA), considering robust SE to adjust for the cluster structure (school level). Covariates such as age, gender, intelligence quotient, auditory processing, and socioeconomic status were considered in the regression models. When intelligence quotient was used as the outcome, it was excluded from the covariates selection.

Results

IRT

Tables 1 and 2 show the discrimination ('a') and difficulty ('b') parameters of the TDE's subtests (reading and writing, respectively) with their respective SE.

Regarding the reading subtest, 61 of 70 items showed very high discrimination ('a'>1.7; the range of discrimination among all items was from 9.599 to 1.339); in the writing subset, 10 of 34 items showed very high discrimination (the range of discrimination was from 3.720 to 0.524). Therefore, for the next step, LCA, we considered 71 items (ie, 61+10) from the reading and writing domains.

LCA

Table 3 shows the fit indices for each model solution (with two, three, four and five-latent class solutions). The two-class solution has acceptable indices and grouped the children into two latent groups: children with poor decoding and writing skills (18.50%) and good decoding and writing skills (81.50%); however under the two-class solution the AIC, BIC, and ssaBIC are highest (among the other solutions). This sort of categorization (ie, dichotomization) might be helpful when it is important to have highly contrasting groups. However, the best statistical solution (AIC = 76,422.476, BIC = 77,560.866, ssaBIC = 76,934.954, entropy = 0.982) was achieved with the three-class solution, in which each class was labeled as follows: children with good decoding and writing skills (GDW; 63.70% of population), not-so-good decoding and writing skills (NsgDW; 22%), and poor decoding and writing skills (PDW; 14.30%), indicating that there is an intermediate class.

As might be noted from Table 4, the majority of children (403) who were classified as having NsgDW skills in the three-classes solution were categorized as children with GDW

Table 1 Discrimination ('a') and difficulty ('b') parameters from the TDE's reading subtest

Translation to English	Word in Portuguese	Estimate 'a'	SE	Estimate 'b'	SE
Now	Agora	**9.599**	3.474	–0.966	0.303
Window	Janela	**9.174**	5.142	–1.132	0.299
Born	Nascimento	**6.274**	1.301	–0.923	0.303
Word	Palavra	**5.792**	0.685	–0.964	0.294
Duck	Pato	**5.706**	1.903	–1.306	0.325
My (female form)	Minha	**5.186**	1.027	–0.964	0.305
Tape	Fita	**5.043**	1.432	–1.137	0.304
Dear (male form)	Querido	**4.927**	1.171	–0.917	0.306
Coin	Moeda	**4.889**	0.92	–1.009	0.286
Brick	Tijolo	**4.748**	0.934	–1.077	0.295
Bait	Isca	**4.271**	0.836	–1.019	0.281
Project	Projeto	**3.983**	0.615	–0.886	0.296
Size	Tamanho	**3.965**	0.845	–0.857	0.295
Field	Campo	**3.922**	0.811	–0.885	0.286
Globe	Globo	**3.801**	0.756	–0.796	0.3
Truck	Caminão	**3.725**	0.567	–0.933	0.289
Success	Sucesso	**3.608**	0.566	–0.848	0.303
Art	Arte	**3.602**	0.735	–0.954	0.274
Shoe	Sapato	**3.589**	0.925	–1.164	0.302
Needle	Agulha	**3.524**	0.568	–0.841	0.296
Storm	Tempestade	**3.399**	0.711	–0.823	0.282
Brushwood	Mato	**3.378**	0.818	–1.229	0.312
Armor	Armadura	**3.348**	0.64	–0.854	0.278
Clover	Trevo	**3.283**	0.628	–0.824	0.276
Tray	Bandeja	**3.282**	0.801	–0.849	0.286
Florist	Floresta	**3.173**	0.655	–0.794	0.291
Brute, inhuman person	Bruto	**3.112**	0.483	–0.817	0.287
Candle	Vela	**3.103**	0.591	–1.142	0.303
Cashew	Caju	**2.954**	0.601	–1.094	0.303
Bone	Osso	**2.930**	0.6	–1.093	0.29
Rapidity	Rapidez	**2.913**	0.576	–0.782	0.287
Cigarette lighter	Isqueiro	**2.890**	0.615	–0.776	0.287
Wolf	Lobo	**2.861**	0.541	–1.080	0.312
Honey	Mel	**2.813**	0.395	–0.786	0.286
Guitar	Guitarra	**2.791**	0.482	–0.561	0.275
Diligent	Aplicado	**2.707**	0.468	–0.735	0.268
Lawyer	Advogado	**2.646**	0.421	–0.61	0.294
Sack	Saco	**2.595**	0.565	–1.146	0.301
Sourness	Azedo	**2.585**	0.392	–0.874	0.269
Fear	Medo	**2.584**	0.479	–1.121	0.294
Garage	Garagem	**2.570**	0.546	–0.745	0.288
1. To abuse, 2. To go beyond limits or measure	Abusar	**2.390**	0.462	–0.621	0.264
Kiosk	Quiosque	**2.385**	0.452	–0.665	0.277
Shaker	Chocalho	**2.349**	0.474	–0.674	0.273
Explanation	Explicação	**2.221**	0.44	–0.604	0.276
Sheet	Lençóis	**2.142**	0.368	–0.557	0.252
Luxurious	Luxuoso	**2.119**	0.323	–0.602	0.266
Aeronautics	Aeronáutica	**2.076**	0.316	–0.549	0.256
Atmosphere	Atmosfera	**2.074**	0.32	–0.354	0.281
Claw	Garra	**2.055**	0.285	–0.726	0.261
Atlas	Atlas	**1.979**	0.319	–0.43	0.264
Dripped (past tense of 'to drip')	Pingado	**1.918**	0.345	–0.593	0.252
Lodging house	Hospedaria	**1.905**	0.348	–0.518	0.248

(*Continued*)

Table 1 (*Continued*)

Translation to English	Word in Portuguese	Estimate 'a'	SE	Estimate 'b'	SE
Scotch tape	Durex	**1.889**	0.29	–0.187	0.257
Exhausted	Exausto	**1.877**	0.284	–0.291	0.259
Curdled milk	Coalhada	**1.872**	0.346	–0.492	0.242
Fatty part of milk, butterfat	Nata	**1.786**	0.362	–1.029	0.285
Repugnant	Repugnante	**1.775**	0.269	–0.092	0.26
Backside	Costas	**1.772**	0.277	–0.618	0.242
Brought	Trouxe	**1.758**	0.275	–0.325	0.25
To get up	Acordar	**1.735**	0.321	–0.66	0.226
Downcast	Acabrunhado	1.624	0.318	–0.385	0.234
Perseverance	Perseverança	1.619	0.263	–0.265	0.225
Exceptional	Excepcional	1.537	0.231	0.062	0.227
Rescinded (past tense of 'to rescind')	Rescindido	1.492	0.269	–0.393	0.234
To ricochet	Ricochetear	1.487	0.213	–0.123	0.2
Hall	Saguões	1.483	0.227	–0.116	0.243
Marsupials	Marsupiais	1.417	0.247	–0.386	0.229
Hypocrite	Hipócrita	1.409	0.212	0.042	0.217
To pique	Vangloriar	1.339	0.183	0.062	0.185

Note: Only values in bold were used to conduct LCA.
Abbreviations: LCA, latent class analysis; SE, standard error; TDE, Academic Performance Test.

skills in the two-class solution. The four-class solution does not have good interpretability either in terms of reading and writing skills or statistics.

Discriminant and concurrent validation

According to Table 5, the three-class solution was not found to predict a major depression diagnostic (ie, having NsgDW skills [odds ratio {OR}: 1.373, $P = 0.315$] or PDW skills [OR: 1.534, $P = 0.424$] in relation to GDW skills), indicating a discriminant validation; in other words, our latent groups were not correlated with depression diagnosis.

Regarding the concurrent validity, children with NsgDW and PDW skills were shown to have approximately twice the probability of combined ADHD in relation to children with GDW skills (OR: 2.056, $P = 0.015$ and OR: 2.258, $P < 0.027$, respectively).

In the same way, multinomial logistic regressions showed in Table 6, where the estimated IQs were tested as predictors of the three-class solution's outcomes, considering both Brazilian norms and the Tellegen and Briggs method, returned the result that NsgDW and PDW skills were predictors of estimated IQ, indicating concurrent validity.

Discussion

We were able to demonstrate that a three-class solution encompassing children with good, not-so-good, and poor decoding (ie, word-level reading) and writing performance presented the best fit for our data. We also showed that this three-class latent solution is valid and reliable because it is clearly associated with ADHD and IQ and it is not associated with emotional disorders such as depression.

It is important to stress, regarding IRT results, due to wide age range, the interpretation of IRT's parameters (ie, 'a' and 'b') will give values of discrimination and difficulty centered on the mean age (9.52 years old, SD = 1.85). It is expected that the items (for reading and writing subtests) become easier throughout the children's development process and therefore the values for 'b' will decrease significantly; this effect is called the schooling effect, and it is expected due to the children's increasing literacy. Discrimination concerning items also can change throughout a student's school years; however, it is more stable regarding the schooling effect than difficulty parameter (ie, 'b'), the higher it is ('a'), the higher the information that 'a' brings to the latent trait (ie, θ); therefore, a very high 'a' was chosen for the cut-off parameter.

The majority of the writing subtest was not used in the LCA, because they scored below the Baker's cut-off parameter for a very high discriminative item (according to Baker, a high discriminative score is 1.35–1.69, moderate is 0.65–1.34, low is 0.35–0.64, and very low is 0.01–0.34).[17] However, we cannot say that they did not contribute to measuring the underlying construct of writing ability, because all writing and reading items showed a factor loading statistically significant due to the fact that *t*-ratios for each item were ≥1.96 (ie, indicating that the factor loading had a *P*-value < 0.05. This can be obtained by dividing the 'a' estimate by its SE. For example, taking the last three

Table 2 Discrimination ('a') and difficulty ('b') parameters from the TDE's writing subtest

Translation to English	Word in Portuguese	Estimate 'a'	SE	Estimate 'b'	SE
To see	Ver	**3.720**	1.080	–0.878	0.266
More	Mais	**2.561**	0.600	–0.925	0.313
Only	Apenas	**2.520**	0.578	–0.895	0.279
Burrow	Toca	**2.520**	0.292	–1.226	0.385
Hammer blow	Martelada	**2.415**	0.379	–0.658	0.237
Kindness	Favor	**2.053**	0.35	–0.576	0.247
Substantive of breaking	Quebramento	**2.026**	0.273	–0.379	0.255
Collectivity	Coletividade	**1.953**	0.308	–0.577	0.267
Unknown	Desconhecido	**1.862**	0.258	–0.107	0.289
Balance	Balance	**1.844**	0.262	–0.084	0.267
Effective	Efetivo	1.657	0.267	–0.672	0.269
Fortification	Fortificação	1.634	0.213	0.049	0.258
Lugubrious	Soturno	1.584	0.175	–0.264	0.24
To crystallize	Cristalizar	1.536	0.201	0.23	0.276
Ball	Baile	1.513	0.219	–0.502	0.256
Prestigious	Prestigioso	1.468	0.23	0.221	0.265
Tap	Bica	1.455	0.304	–1.072	0.332
To digest	Digerir	1.451	0.197	–0.116	0.251
Discriminative	Discriminativo	1.445	0.219	0.189	0.245
Boisterous	Revolta	1.403	0.213	0.100	0.248
To commercialize	Comercializar	1.365	0.159	0.769	0.25
To bring before the court	Ajuizar	1.333	0.159	0.41	0.245
Laziness	Preguiça	1.313	0.143	0.823	0.221
To take the lid off	Destampar	1.305	0.161	0.419	0.228
Industrialization	Industrialização	1.226	0.147	0.867	0.264
Composition	Composição	1.194	0.172	0.463	0.253
Shivering sensation	Calafrio	1.168	0.174	0.146	0.254
Legitimacy	Legitimidade	1.122	0.151	0.282	0.22
Helmet	Elmo	1.120	0.16	0.345	0.225
To be comforted	Consolado	1.092	0.17	0.079	0.276
Impetuosity	Impetuosidade	1.019	0.095	1.295	0.23
Manly	Varonil	1.011	0.147	0.208	0.236
Similarity	Similaridade	1.010	0.152	0.55	0.211
Quick	Rápida	0.524	0.063	0.923	0.208

Note: Only values in bold were used to conduct LCA.
Abbreviations: LCA, latent class analysis; SE, standard error; TDE, Academic Performance Test.

discrimination estimates from the writing subtest, we have the following *t*-ratios: for 'varonil' [discrimination estimate/SE = 1.011/0.147 = 6.877], 'similaridade' [discrimination estimate/SE = 1.010/0.152 = 6.64], and 'e rápida' [discrimination estimate/SE = 0.524/0.063 = 8.317]).

The bootstrapped likelihood ratio test did not help in deciding on the number of latent classes, because its *P*-values were <0.0001 for any reasonable number of latent classes. One explanation may be that the 71 items were not designed to capture latent classes; as described, they were selected based on the discrimination parameter supplied by IRT. In contrast, the three-class solution model may be a good enough approximate model and does give a reasonable interpretation of reading and writing skills. However, because the *P*-value

Table 3 LCA fit indices fit for the two-, three-, and four-class solutions

N classes	Free parameters	H0	AIC	BIC	ssaBIC	Entropy	LMR ratio	*P*-value	Bootstrapped likelihood ratio test
2	140	–46537.604	93355.209	94157.679	93712.874	0.998	*	*	*
3	197	–38014.238	76422.476	77560.866	76934.954	0.982	–44414.643	<0.0001	<0.0001
4	263	–36148.717	72823.433	74343.213	73507.605	0.968	–38014.238	0.3323	<0.0001
5	329	–35051.659	70761.318	72622.488	71617.183	0.947	–36148.717	0.1092	<0.0001

Note: *No value.
Abbreviations: AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; ssaBIC, sample size adjusted; LMR, Lo–Mendell–Rubin likelihood ratio; LCA, latent class analysis; H0, null hypothesis.

Table 4 Cross-tabs of two- and three-class solutions

Two-class solution	Three-class solution			Total
	GDW skilled	NsgDW skilled	PDW skilled	
GDW skilled	1,238	348	0	1,586
PDW skilled	0	80	279	359
Total	1,238	428	279	1,945

Abbreviations: GDW, good decoding and writing skills; NsgDW, not-so-good decoding and writing skills; PDW, poor decoding and writing skills.

derived from the Lo–Mendell–Rubin test is higher than 0.05 ($P = 0.323$), the three-class model was not rejected when compared to the four-class solution, corroborating the previous findings from other fit indices.

The class counts based on the most likely posterior class for the four-class solution were, in terms of percentage: 58.62% (related to GDW skilled children), 22.74% and 7.65% (related to two splits of the intermediate category [ie, NsgDW skilled children], which was observed between GDW and PDW skilled children), and 10.98% related to PDW. As can be seen, the two intermediate latent groups within the four-class solution, in terms of reading and writing abilities, do not have a fair interpretation. Regarding statistics, the sample proportion (and as a consequence, the size) of each class within the four-class solution is too small, which also does not help much in the interpretability of classes. The reduction in the BIC and ssaBIC was about 3,000. However, taking into account the reduction from the two- to the three-class solution, the decline was around 16,000 for AIC, BIC, and ssaBIC; the progression in the reduction for the five-class solution is still around 2,000 (ie, from the four-class solution to the five-class solution, the difference is: 2,062 for AIC, 1,721 for BIC, and 1,890 for ssaBIC).

The four-class solution does not have good interpretability either in terms of reading and writing skills or statistics due to the low proportion within one of the intermediate latent classes (7.56%). The three-class solution is preferable due to previous evidence and due to the nature of TDE's items (basically centered on the decoding skills and, therefore, any other reading/writing domains such as comprehension). Within the four-class solution, two intermediate latent classes emerged as not having any atypical profile (such as high reading but low writing skills or vice versa); we opted, therefore, for the more parsimonious structure (GDW, NsgDW, and PDW skilled children). Indeed, regarding the number of underlying latent classes exclusively related to reading skills, a recent study has identified a three-class solution using 27 items (17 items pertaining to children's reading aloud and ten to silent reading) from the Scale of Evaluation of Reading Competence of Students by the Teacher (EACOL). The EACOL was used to evaluate reading aloud and silent reading among a heterogenic sample of elementary school children (N = 335 children, with an average age of 9.75 years [SD = 1.2]); the classes were called good reader, not-so-good reader, and poor reader.[24]

No relation was observed between the three-class solution and a DAWBA diagnostic of major depression. Because the underlying constructs of the three-class solution were based on items related to decoding and writing skills, no correlation was expected to be found between latent classes and major depression. Our finding is in accordance with a study that found that reading difficulty did not directly influence the level of self-evaluation or depression.[25] It is important to stress that a meta-analysis quantified the mean differences in depression measure scores and levels of clinical depression

Table 5 Discriminating validity of three-class solution through logistic regression

Outcomes	Tested predictor	Estimate (odds ratio)	Robust SE	*P*-value	95% confidence interval	
Major depression diagnostic (binary outcome)	NsgDW skills group	1.447	0.472	0.258	0.762	2.744
	PDW skills group	1.721	0.803	0.244	0.689	4.299
	Age	1.296	0.1058	0.001	1.104	1.521
	Female	1.133	0.332	0.670	0.637	2.015
	SES	0.652	0.165	0.092	0.397	1.072
	IQ	1.009	0.008	0.261	0.992	1.026
	Problems on hearing	1.396	0.282	0.098	0.939	2.075
Combined ADHD diagnostic (binary outcome)	NsgDW skills group	2.056	0.608	0.015	1.151	3.673
	PDW skills group	2.258	0.834	0.027	1.837	4.659
	Age	0.966	0.768	0.665	0.826	1.129
	Female	0.536	0.149	0.026	0.310	0.927
	SES	0.800	0.178	0.319	0.517	1.239
	IQ	0.983	0.608	0.016	0.970	0.996

Abbreviations: ADHD, attention deficit hyperactivity disorder; NsgDW, not-so-good decoding and writing skills; PDW, poor decoding and writing skills; SES, socioeconomic status; SE, standard error; IQ, intelligence quotient.

Table 6 Concurrent validity of estimated IQ (Brazilian normalization) and Tellegen and Briggs methods through multinominal regression

Outcomes	Tested predictor	Estimate	Robust SE	*P*-value	Confidence interval 95%	
NsgDW skills group	Estimated IQ/Brazilian normalization	–0.331	0.042	<0.001	–0.414	–0.248
	Age	–0.554	0.047	<0.001	–0.646	–0.462
	Female	–0.293	0.095	0.001	–0.472	–0.114
	SES	–0.074	0.095	0.434	–0.260	0.111
	Problems on hearing	–0.233	0.131	0.076	–0.491	0.024
PDW skills group	Estimated IQ/Brazilian normalization	–0.576	0.069	<0.001	–0.712	–0.440
	Age	–1.304	0.840	<0.001	–1.469	–1.140
	Female	–0.531	0.205	0.010	–0.933	–0.128
	SES	–0.383	0.175	0.028	–0.726	–0.040
	Problems on hearing	–0.033	0.171	0.847	–0.369	0.303
NsgDW skills group	Estimated IQ (Tellegen and Briggs[11])	–0.030	0.003	<0.001	–0.377	–0.022
	Age	–0.546	0.046	<0.001	–0.636	–0.455
	Female	–0.286	0.091	0.002	–0.464	–0.455
	SES	–0.080	0.094	0.392	–0.266	0.104
	Problems on hearing	–0.221	0.123	0.086	–0.473	0.031
PDW skills group	Estimated IQ (Tellegen and Briggs[11])	–0.049	0.006	<0.001	–0.062	–0.371
	Age	–0.576	0.069	<0.001	–1.461	–1.139
	Female	–0.513	0.205	0.013	–0.917	–0.110
	SES	–0.389	0.177	0.028	–0.737	–0.041
	Problems on hearing	–0.017	0.170	0.921	–0.351	0.317

Note: The presented estimations for each latent class (NsgDW and PDW) were the results from each one compared to the reference value (GDW), which was omitted due the statistical comparison.

Abbreviations: ADHD, attention deficit hyperactivity disorder; GDW, good decoding and writing skills; NsgDW, not-so-good decoding and writing skills; PDW, poor decoding and writing skills; SES, socioeconomic status; SE, standard error; IQ, intelligence quotient.

between students with and without learning disabilities, and it concluded that there is a lack of data-based studies on depression among students with learning disabilities.[26] However, it must be emphasized that in this study, it was not possible to measure learning disorders/disabilities, due to the nature underlying the 71 items that formed the latent classes. Regarding the concurrent validity of IQ as a predictor of the three-class solutions' outcomes (the children with GDW skills as the reference group in a multinomial regression model), children with NsgDW and PDW skills were predicted by both estimated IQ measures. The lower the class (ie, PDW), the lower the estimated IQ score, with GDW-skilled children constituting the reference group. In other words, IQ was shown to be a predictor of latent classes. A possible explanation is that, because estimated IQ is calculated using vocabulary and cubes subtests, some studies have pointed out that semantic knowledge is important for recognizing individual words.[27,28] Moreover, the results are in accordance with a study that found that correlations between reading achievement tests and Weschler intelligence tests were substantially higher when the intelligence tests evaluated verbal abilities than when they evaluated non-verbal abilities. The magnitude of correlation between verbal IQ and word identification was around 0.62 (which is considered moderate) among older readers[29] (older readers, in this paper, were understood to be in the 6th and 7th grades). It should be mentioned that this finding suggests that correlations between reading achievement and tests of intelligence may be an artifact of shared variance to which language-based abilities underlying performance on both sets of measures contribute.[30] A recent publication highlighted the major concerns and controversies surrounding the *Diagnostic and Statistical Manual of Mental Disorders IV* for learning disorders; a concern with the IQ-achievement discrepancy criterion, for specific learning disorders, stated that the estimation of discrepancy is statistically flawed.[31] Based on convergent evidence from the literature review, the above cited work group recommended that IQ-achievement discrepancy criterions used in DSM-IV should be eliminated.

Corroborating the findings about the concurrent validity of the three-class solution and combined ADHD diagnostic, the authors have pointed out the frequent co-occurrence of ADHD and reading disability: 25%–40% of individuals with one disorder also meet the diagnostic criteria for the other.[32–34] Considering reading disability and ADHD symptoms (inattentiveness and hyperactivity-impulsivity) as contiguous

traits in population samples, other evidence about the correlation between the two might be pointed out.[35–37] In the same way, a research study using a larger sample of 15-year-old children from public schools examined psychiatric morbidity and functional impairment of adolescents with and without poor reading skills during mid to late adolescence. Among children with poor reading skills, it found higher rates of concurrent attention deficit/hyperactivity, affective, and anxiety disorders, particularly social phobia and generalized anxiety disorder.[38]

Regarding the covariate age used on models where estimated IQs (the three-class solution was used as an outcome to test the estimated IQs' prediction) were considered as an outcome, it is important to point out that a schooling effect was observed in which the higher the education level, the better the achievement in the accuracy of words.[24] This finding strengthens the findings about the three-class solution's concurrent validity. Also, in this same modelings' contexts (ie, the three-class solution as an outcome), SES appeared to be a predictor of the lower classes (PDW skills) in comparison to the higher level class (GDW skills). This is consistent with previous studies where various developmental antecedents (eg, social deprivation, socioeconomic status, family size, maternal reading, a stimulating home environment, maternal depression, and child negligence) have been shown to have a small but significant relation to reading achievement.[8,39]

One strength to be addressed in this study is the possibility of associating IRT and LCA in order to provide, based on very highly discriminative items ('a'>1.7), a categorical solution for a construct, which might be thought of as offering a person-centered perspective. One limitation, which is inherent to this type of analysis, is that this LCA is restricted to the given sample (ie, 1,945 subjects and their achievements and subgroups). However, it is important to stress that the three-class solution for decoding and writing skills was shown to be reliable and valid. The selection of the highly discriminative items via IRT involving differential item functioning between genders or among different school regions (ie, São Paulo/Porto Alegre) was not considered for discussion here. In other words, we did not test to see if examinees with equal ability but who are from different, unequal groups have an unequal probability of item success, because this paradigm is directly related to a dimensional view of the reading/writing construct. Therefore, it will be important to test differential item functioning in future research exploring the possibility of a dimensional construct underlying the TDE; this could be explored using the factor analysis with covariates or the multiple-group approach.

Conclusion

The three-class solution was shown to have discriminant and concurrent validity, enabling a more accurate grouping of students according to good, not-so-good, or poor performance and making this set of items a reliable indicator of deficits. The latent classes can assist researchers who require reliable and validated categories for diagnosis or identification of other behavior and performance patterns.

Disclosure

The authors report no conflicts of interest.

References

1. Pinheiro A, Rothe-Neves R. Avaliação cognitiva de leitura e escrita: as tarefas de leitura em voz alta e ditado [Cognitive assessment of reading and writing: reading aloud and spelling tasks]. *Psicol Refl Crít*. 2001;14:399–408. Brazilian Portuguese.
2. Stein LM. TDE Teste de desempenho escolar [Academic Performance Test]. São Paulo: Casa do Psicológo; 1998. Brazilian Portuguese.
3. Lúcio PS, Pinheiro A, Nascimento E. O impacto da mudança no critério de acerto na distribuição dos escores do subteste de leitura do Teste de Desempenho Escolar [The impact of the change in the criterion of scoring correct responses in the distribution of the scores of the reading subtest of performance school test]. *Psicologia em Estudo*. 2009;14:593–601. Brazilian Portuguese.
4. Lazarsfeld PF, Henry NW. *Latent Structure Analysis*. Boston: Houghton Mifflin; 1968.
5. Clogg C. *Latent Class Models*. New York: Plenum; 1995.
6. Everitt B, Hand D. Finite mixture distribution. In: Cox DR, Isham N, Keiding T, et al, editors. *Monographs on Applied Probability and Statistics*. London: Chapman and Hall; 1981.
7. Magidson J, Vermunt J. Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*. 2002;20:37–44.
8. Pepe MS, Janes H. Insights into latent class analysis of diagnostic test performance. *Biostatistics*. 2007;8:474–484.
9. Salum GA, Mogg K, Bradley BP, et al. Threat bias in attention orienting: Evidence of specificity in a large community-based study. *Psychol Med*. 2013;43:733–745.
10. Wechsler D. *WISC-III: Escala de Inteligência Wechsler para Crianças: Manual [Wechsler Intelligence Scale for Children: Manual]*. São Paulo: Casa do Psicólogo; 2002. Brazilian Portuguese.
11. Tellegen A, Briggs PF. Old wine in new skins: grouping Wechsler subtests into new scales. *J Consult Psychol*. 1967;31:499–506.
12. Figueiredo VLM. *Uma adaptação brasileira do teste de inteligência WISC-III [A Brazilian adaptation of Wechsler Intelligence Scale for Children (WISC-III)]*. Brasília, DF: Curso de Pós-Graduação em Psicologia, Instituto de Psicologia – Universidade de Brasília; 2001. Brazilian Portuguese.
13. Engelmann L, Ferreira M. Avaliação do processamento auditivo em crianças com dificuldades de aprendizagem. [Assessment of hearing processing among children with learning difficulties]. *Rev Soc Bras Fonoaudiol*. 2009;14(1):69–74. Brazilian Portuguese.
14. Banai K, Kraus N. Neurobiology of (central) auditory processing disorder and language-based learning disabilities. In: Chermak G, Musiek F, editors. *Handbook of Central Auditory Processing Disorder*. San Diego: Publishing Ink; 2006:89–199.
15. ABEP. Critério de classificação econômica Brasil [Brazilian economic criteria classification]. Associação Brasiliera de Empresas de Pesquisa (ABEP); 2010.

16. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The development and well-being assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry*. 2000;41:645–655.
17. Pasquali L. TRI – Teoria de Resposta ao Item: Teoria, Procedimentos e Aplicações [IRT – Item Response Theory? Theory, Procedures, and Application]. Vol 1. Brasília, DF: Laboratório de Pesquisa em Avaliação e Medida – LabPAM; 2007:2001.
18. Baker F. *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation; 2001.
19. Muthén B. Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In: Kaplan D, editor. *Handbook of Quantitative Methodology for the Social Sciences*. Newbury Park: Sage; 2004:345–368.
20. Lo Y, Mendell N, Rubin D. Testing the number of components in a normal mixture. *Biometrika*. 2001;88:767–778.
21. Vuong Q. Likelihood ratio test for model selection and non-nested hypotheses. *Econometrica*. 1989;57:307–333.
22. Asparouhov T, Muthén B. Using Mplus TECH11 and TECH14 to test the number of latent classes. Mplus Web Notes: No 14, May 22, 2012. Available from: http://www.statmodel.com/examples/webnotes/webnote14.pdf. Accessed May 3, 2013.
23. Muthén L, Muthén B. *Mplus Users' Guide*. 4th ed. Los Angeles: Muthén and Muthén; 2010.
24. Cogo-Moreira H, Ploubidis GB, de Avila CR, de Jesus Mari J, Pinheiro AM. EACOL (Scale of Evaluation of Reading Competence by the Teacher): Evidence of concurrent and discriminant validity. *Neuropsychiatr Dis Treat*. 2012;8:443–454.
25. Yamashita T, Hayashi T. Reading ability of junior high school students in relation to self-evaluation and depression. *No To Hattatsu*. 2012;44:45–49. Japanese.
26. Maag JW, Reid R. Depression among students with learning disabilities: Assessing the risks. *J Leanr Disabil*. 2006;39:3–10.
27. Nation K, Snowling MJ. Semantic processing and the development of word recognition skills: Evidence from children with reading comprehension difficulties. *J Mem Lang*. 1998:85–101.
28. Plaut DC, McClelland JL, Seidenberg MS, Patterson K. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychol Rev*. 1996;103:56–115.
29. Vellutino FR, Scanlon DM, Tanzman MS. Components of reading ability: Issues and problems in operationalizing word identification, phonological coding, and orthographic coding. In: Lyon GR, editor. *Frames of Reference for the Assessment of Learning Disabilities: New Views on Measurement Issues*. Baltimore: Brookes; 1994: 279–324.
30. Vellutino FR, Scanlon DM, Lyon GR. Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. *J Learn Disabil*. 2000;33:223–238.
31. Tannock R. Rethinking ADHD and LD in DSM-5: Proposed changes in Diagnostic Criteria. *J Learn Disabil*. 2013;46(1):5–25.
32. August GJ, Garfinkel BD. Comorbidity of ADHD and reading disability among clinic-referred children. *J Abnorm Child Psychol*. 1990;18:29–45.
33. Willcutt EG, Pennington BF. Psychiatric comorbidity in children and adolescents with reading disability. *J Child Psychol Psychiatry*. 2000;41:1039–1048.
34. Semrud-Clikeman M, Biederman J, Sprich-Buckminster S, Lehman BK, Faraone SV, Norman D. Comorbidity between ADHD and learning disability: A review and report in a clinically referred sample. *J Am Acad Child Adolesc Psychiatry*. 1992;31:439–448.
35. Stevenson J, Langley K, Pay H, et al. Attention deficit hyperactivity disorder with reading disabilities: Preliminary genetic findings on the involvement of the ADRA2A gene. *J Child Psychol Psychiatry*. 2005;46:1081–1088.
36. Willcutt EG, Pennington BF, Olson RK, DeFries JC. Understanding comorbidity: A twin study of reading disability and attention-deficit/hyperactivity disorder. *Am J Med Gent B Neuropsychiatric Genet*. 2007;144B:709–714.
37. Paloyelis Y, Rijsdijk F, Wood AC, Asherson P, Kuntsi J. The genetic association between ADHD symptoms and reading difficulties: The role of inattentiveness and IQ. *J Abnorm Child Psychol*. 2010;38:1083–1095.
38. Goldston DB, Walsh A, Mayfield Arnold E, et al. Reading problems, psychiatric disorders, and functional impairment from mid- to late adolescence. *Am Acad Child Adolesc Psychiatry*. 2007;46(1):25–32.
39. Trzesniewski KH, Moffitt TE, Caspi A, Taylor A, Maughan B. Revisiting the association between reading achievement and antisocial behavior: New evidence of an environmental explanation from a twin study. *Child Dev*. 2006;77:72–88.