

# Desenvolvimento do subteste de escrita do Teste de Desempenho Escolar II

Marcia de Lima Athayde<sup>1</sup>, Claudia Hofheinz Giacomoni, Euclides José de Mendonça Filho

Universidade Federal do Rio Grande do Sul, Porto Alegre-RS, Brasil

Rochele Paz Fonseca, Lilian Milnitsky Stein

Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre-RS, Brasil

## RESUMO

Este estudo teve como objetivo o desenvolvimento do subteste de escrita da segunda edição do Teste de Desempenho Escolar (TDE II). Ele envolveu a construção de itens, a análise de juízes e uma testagem empírica. Contou com a participação de 684 estudantes de 1º a 9º ano do ensino fundamental de escolas públicas e privadas da região metropolitana de Porto Alegre (RS). O subteste foi dividido em duas versões: para estudantes de 1º a 4º ano e para de 5º a 9º ano. A análise fatorial apontou um fator explicando a variância comum dos itens, sendo um fator explicando 71% na versão A e um fator explicando 56% na versão B. A Teoria de Resposta ao Item (TRI) mostrou níveis de discriminação e de dificuldade dos itens adequados para mensurar o construto. Os resultados revelaram evidências de validade de conteúdo, construto e fidedignidade. O subteste de escrita apresentou-se com boas propriedades psicométricas para avaliação da escrita ao longo do ensino fundamental.

*Palavras-chave:* rendimento escolar; escrita; psicomетria; Teoria de Resposta ao Item.

## ABSTRACT – Development of the writing subtest of the School Performance Test II

This study aimed to develop the new version of the writing subtest of School Performance Test – second edition (TDE II). It consisted in creating items, judge analysis and empirical testing. It had the participation of 684 students from the first to the ninth grade of elementary school of public and private institutions from Porto Alegre (RS), Brazil. The subtest was divided in two versions: one for students from the first to the fourth grade and one for students from the fifth to the ninth grade. Factor analysis indicated a factor explaining the common variance of the items, with one factor explaining 71% in version A and one factor explaining 56% in version B. The Item Response Theory (TRI) showed appropriate levels of discrimination and difficulty, which suggested that the items provided a comprehensive measure of the latent trait. The results provided satisfactory evidence of internal validity and reliability. These findings showed that the writing subtest presents good psychometric properties for the writing performance evaluation through elementary school.

*Keywords:* academic achievement; writing; psychometrics; Item Response Theory.

## RESUMEN – Desarrollo de la sub-prueba de escritura de la Prueba del Desempeño Escolar II

Este estudio tuvo como objetivo desarrollar la sub-prueba de escritura de la segunda edición de la Prueba del Desempeño Escolar (TDE II). El estudio involucró en la construcción de ítems, el análisis de los jueces y una comprobación empírica. Se contó con la participación de 684 alumnos del primero a noveno grado de la escuela primaria, de escuelas públicas y privadas, en la región metropolitana de Porto Alegre (estado del Rio Grande do Sul, Brasil). La sub-prueba se dividió en dos versiones: para los estudiantes de primer al cuarto grado y para los de quinto a noveno grado. El análisis factorial indicó un factor que explica el 71,0 (versión A) y 56,0% (versión B) de la varianza de los datos. La Teoría de la Respuesta al Ítem (TRI) mostró niveles de discriminación y dificultad de los elementos adecuados para medir el constructo. Los resultados mostraron evidencia de la validez del contenido, constructo y fiabilidad. La sub-prueba de escritura se presenta con buenas propiedades psicométricas para evaluación de la escritura en toda la escuela primaria.

*Palabras clave:* rendimiento escolar; escritura; psicomетría; Teoría de Respuesta al Ítem.

O ensino da habilidade da escrita é fundamental, bem como sua avaliação em cada ano escolar. A escrita envolve diversos processos interdependentes, como a capacidade de conversão fonema-grafema, a ortografia e a escrita de frases e textos. A escrita de palavras precede a escrita de frases e textos, embora não garanta uma escrita

de textos adequada. Um aluno pode não ter competência para escrever um texto apropriado, não por conta da dificuldade de produção textual, mas sim porque não domina a conversão fonema-grafema.

O desenvolvimento da escrita começa nos anos pré-escolares e continua ao longo dos anos de escolarização,

<sup>1</sup>Endereço para correspondência: Rua Ramiro Barcelos, 2.492, Santa Cecília, 90035-003, Porto Alegre-RS, Brasil. E-mail: marcialadoc@gmail.com

possivelmente estendendo-se até a idade adulta (Costa, Edwards, & Hooper, 2016). Inicialmente, a aprendizagem da escrita começa com o domínio de produzir letras legíveis e a ortografia básica (Abbott, Berninger, & Fayol, 2010). Uma vez que essas competências são obtidas, passa-se para níveis mais avançados da aprendizagem, como o domínio da gramática, a estrutura das sentenças e demais aspectos textuais (Crossley, Weston, & Sullivan, 2011).

Em vista disso, é necessário que existam instrumentos de avaliação variados, que contemplem alunos com diferentes níveis de conhecimento sobre a escrita (Ferreira & Leal, 2007). Em relação à avaliação da escrita no cenário brasileiro, há tarefas envolvendo a avaliação da escrita de palavras (Capovilla & Capovilla, 2003; Moojen, 2009), escrita de palavras e pseudopalavras (Capovilla, Smythe, Capovilla, & Everatt, 2001; Pinheiro, 1995; Salles, 2005; Salles & Parente, 2007) e avaliação da ortografia (Batista & Capellini, 2011).

No que tange à avaliação do desempenho escolar, há uma importante demanda de instrumentos com esse fim no Brasil. O Teste de Desempenho Escolar (TDE) (Stein, 1994) é um instrumento de triagem destinado à avaliação ampla da aprendizagem. É composto por três subtestes que avaliam as capacidades básicas envolvidas no desempenho escolar: leitura, escrita e aritmética (Stein, 1994).

O TDE foi criado há 21 anos, o que justifica sua atualização de normas e conteúdo. A revisão periódica de testes é sugerida pela Resolução n.º 02/2003 do Conselho Federal de Psicologia (2003), pela International Test Commission (2000) e pelos padrões para testes psicológicos e educacionais da Associação Americana de Psicologia (APA). Outros fatores corroboram a necessidade de desenvolvimento do TDE II:

1. o TDE avalia sujeitos apenas até a 6ª série (Stein, 1994), não abarcando todo o ensino fundamental;
2. o currículo do ensino fundamental foi modificado, de forma que seus conteúdos foram redistribuídos em nove anos escolares (modificação prevista na Lei n.º 10.172, de 9 de janeiro de 2001);
3. já foi comprovada a larga utilização do TDE tanto para a prática clínica como para o uso em pesquisas científicas nacionais (Knijnik, Giacomoni, & Stein, 2013);
4. o TDE é o único instrumento psicopedagógico validado e normatizado para a população brasileira (Knijnik, Zanon, Giacomoni, & Stein, 2014);
5. especificamente em relação ao subteste de escrita original do TDE, ele apresenta evidências de validade, mas possui lacunas em suas características psicométricas (Giacomoni, Athayde, Zanon, & Stein, 2015).

O subteste de escrita do TDE avalia as capacidades básicas envolvidas na habilidade de escrita: a conversão fonema-grafema e a ortografia. Segundo Conrad, Harris, & Williams (2013), a ortografia é a compreensão das convenções utilizadas em um sistema de escrita, ou o conhecimento de como as palavras são escritas. Em suma,

o subteste analisa a capacidade de o estudante escrever palavras ortograficamente corretas, identificando seu desempenho no nível da palavra. O instrumento não examina os demais aspectos envolvidos na escrita, tal como a escrita de textos. Consiste em um ditado de 34 palavras contextualizadas e na escrita do próprio nome.

É fundamental examinar o nível de competência do aprendiz na habilidade de escrita de palavras, a fim de verificar se está adequado para a idade e o ano escolar. Somente com o desenvolvimento da habilidade de escrita de palavras é que o estudante conseguirá adquirir a capacidade de escrever frases e textos. Portanto, a avaliação da habilidade de codificação da linguagem oral possibilita direcionar o aluno que apresenta dificuldade para avaliações mais abrangentes do referido construto. Além disso, a constatação do nível de habilidade dos estudantes, bem como a identificação de eventuais prejuízos nessa área, viabiliza a criação de métodos de ensino que favoreçam o desenvolvimento da linguagem escrita.

Diante do exposto, o objetivo do presente estudo foi apresentar o processo de desenvolvimento e as evidências de validade de conteúdo do subteste de escrita do TDE II (Stein, Giacomoni, & Fonseca, 2015). Esse processo tomou como base a versão original do subteste de escrita e envolveu a construção de novos itens e a testagem da versão preliminar do subteste em escolares de todo o ensino fundamental I e II (do 1º ao 9º ano).

## Método

O desenvolvimento do subteste de escrita do TDE II foi fundamentado nos procedimentos descritos por Pasquali (2010) para a construção de testes referentes a construtos. Nessa sessão, serão descritos os procedimentos realizados no âmbito dos polos teórico, experimental e analítico. A normatização, última etapa do polo analítico, apresentar-se-á em um estudo futuro.

O polo teórico foi composto pelas etapas de criação dos itens e da análise teórica deles, sendo esta última subdividida em quatro etapas. O polo experimental consistiu na testagem empírica dos itens da versão preliminar do subteste de escrita, já o polo analítico, na análise dos dados obtidos, o qual é exposto dividindo-se os itens do subteste em versões A e B.

### Etapa 1: polo teórico — criação dos itens do subteste de escrita do TDE II

Para a criação dos itens, inicialmente foram definidos critérios empíricos e teóricos para nortear o processo de criação dos itens do subteste de escrita do TDE II. Com base nesses critérios, uma junta de juízes peritos criou o *pool* de itens que fariam parte do subteste.

No que tange aos critérios empíricos, eles foram elegidos tendo em vista os achados experimentais dos itens da versão original do subteste de escrita. Em suma, analisaram-se os dados dos estudos de Knijnik, Zanon,

Giacomoni, e Stein (2014) e Giacomoni et al. (2015), os quais constataram que a grande maioria dos itens do subteste de escrita original apresentava dificuldade média, avaliados por meio da Teoria de Resposta ao Item (TRI). Com isso, o subteste estaria avaliando adequadamente sujeitos com níveis médios de habilidade, mas menos satisfatoriamente estudantes com baixa ou alta habilidade. Fundamentado nesses achados, julgou-se necessário criar itens com níveis de dificuldade variando do fácil ao difícil, visando englobar os nove anos do ensino fundamental.

Após a seleção dos critérios empíricos, uma junta de juízes peritos foi consultada para a definição dos critérios teóricos e a posterior criação dos itens. A junta de juízes foi composta por dois peritos, um com formação em Fonoaudiologia, especialização em Neuropsicologia e em Linguagem e mestrado em Letras (juiz 1), e outro com formação em Pedagogia, Fonoaudiologia e Psicopedagogia, além de mestrado em Educação com ênfase em Psicologia Escolar (juiz 2). Ambos os juízes foram selecionados em função de seu vasto conhecimento teórico e clínico sobre a linguagem escrita.

Os juízes 1 e 2 selecionaram os critérios teóricos com base na literatura relacionada com o tema linguagem escrita. Ainda, com o objetivo de compreender os modos de avaliação da habilidade de escrita de palavras, foram pesquisados outros instrumentos que avaliavam o mesmo construto: ditado balanceado (Moojen, 2009); Escala de Inteligência Wechsler para Crianças (WISC IV), com subtestes de semelhanças e de vocabulário (Rueda, Noronha, Sisto, Santos, & Castro, 2012); Instrumento de Avaliação Neuropsicológica Breve (NEUPSILIN) (Fonseca, Salles, & Parente 2009); Provas de Habilidades Metalinguísticas e de Leitura (PROHMELE) (Cunha & Capellini, 2009); e Provas de Avaliação dos Processos de Leitura (PROLEC) (Capellini, Oliveira, & Cuetos, 2010), além do próprio subteste de escrita original do TDE.

Como resultado das pesquisas realizadas pelos juízes, foram eleitos os seguintes critérios teóricos: nível de aprendizagem de cada ano escolar, uso de conteúdo não regionalizado, termos não pejorativos, ortografia (dificuldades ortográficas) e grau de familiaridade dos itens (palavras). O nível de aprendizagem escolar serviu de base para a criação dos itens, na medida em que foram selecionados itens condizentes com o nível de conhecimento ortográfico de cada ano escolar. Os itens foram concebidos obedecendo ao avanço da aprendizagem da escrita, sendo eles distribuídos nos nove anos escolares do ensino fundamental. O propósito do critério teórico “nível de aprendizagem escolar” era apresentar os itens em ordem crescente de dificuldade.

No que tange à utilização de itens com conteúdo não regionalizado, buscou-se eleger itens que fossem familiares para todo o Brasil, e não apenas para a Região Sul do país, local onde o teste foi criado. Em relação à seleção de termos não pejorativos, optou-se por itens que não

tivessem uma imagem considerada socialmente negativa. Tanto o critério de regionalização quanto o de emprego de termos não pejorativos se mostraram importantes, uma vez que na versão original do TDE havia itens com essas características. No subteste de leitura original havia a palavra *durex*, a qual é conhecida como “fita adesiva” em determinadas regiões do Brasil, bem como o item *isqueiro*, o qual carrega uma imagem socialmente negativa.

Com referência ao critério teórico ortográfico, os peritos 1 e 2 levaram em conta as dificuldades ortográficas da língua portuguesa, bem como a complexidade da estrutura silábica das palavras, fundamentados no trabalho de Moojen (2009). A criação dos itens partiu do pressuposto de que palavras com características ortográficas e estrutura silábica mais simples são mais fáceis para as crianças escreverem. Sendo assim, os juízes criaram para os primeiros anos escolares, por exemplo, itens mais simples, com regras ortográficas e estrutura silábica compatíveis com o ano escolar em questão, já que, segundo Loveall, Channell, Phillips, & Connors (2013), o conhecimento ortográfico é influenciado pela experiência prévia do aprendiz com palavras escritas (impresas). Por exemplo, em relação às regras ortográficas, palavras com dígrafos como *sc*, *ss* e *xc* não foram inseridas até o 3º ano escolar, por conterem ortografia considerada mais complexa para os primeiros anos escolares, de acordo com o parecer dos juízes peritos. Ademais, no que concerne às estruturas silábicas, predominaram nos primeiros anos escolares itens contendo a estrutura CV (consoante seguida de vogal), por serem mais facilmente escritas pelas crianças. No decorrer da escolarização, estruturas silábicas mais complexas (consoante-vogal-consoante — CVC, consoante-vogal-vogal — CVV, consoante-consoante-vogal — CCV) foram inseridas em maior quantidade. Dessa maneira, com o avanço dos anos escolares, a complexidade dos itens acentua-se.

Quanto ao critério teórico chamado de grau de familiaridade das palavras, sabe-se que estímulos familiares no cotidiano das crianças produzem respostas mais rápidas e corretas do que os menos familiares (Cycowicz, Friedman, Rothstein, & Snodgrass, 1997), processo que também ocorre na escrita de palavras (Pinheiro & Rothe-Neves, 2001). Com isso, buscou-se criar itens mais familiares para os primeiros anos escolares e menos familiares para anos mais avançados. Por conseguinte, seria possível aumentar a dificuldade dos itens à medida que os anos escolares avançassem. O julgamento do grau de familiaridade dos itens foi realizado pela junta de peritos mencionada anteriormente, com base em sua experiência na área da linguagem escrita.

Em virtude de os critérios utilizados para a criação do subteste de escrita do TDE II serem distintos dos aplicados no TDE original, optou-se por não manter os itens da primeira versão. Portanto, apenas o item “composição” do subteste de escrita do TDE I permaneceu na versão atualizada, em função de suas características teóricas

serem adequadas aos critérios elegidos para a seleção dos itens (grau de familiaridade, estrutura silábica e dificuldades ortográficas). Dessa forma, ao final da etapa 1, os itens da versão preliminar do subteste de escrita do TDE II haviam sido criados, com base nos critérios teóricos e empíricos supracitados.

## **Etapa 2: polo teórico — análise teórica dos itens do subteste de escrita do TDE-II**

Com a criação dos itens na etapa 1, obteve-se uma versão preliminar do subteste de escrita, a qual continha 118 itens. Com o intuito de verificar evidências de validade de conteúdo, os itens foram analisados por juízes peritos, conforme explicitado a seguir.

### **Análise teórica 1**

Na primeira análise teórica, uma junta de juízes peritos averiguou os itens da versão preliminar do subteste de escrita. Essa junta de juízes foi composta por dois fonoaudiólogos: um com mestrado em Distúrbios da Comunicação Humana (juiz 3); e outro com graduação também em Psicologia, especialista em Linguagem, mestre e doutor em Psicologia do Desenvolvimento e perito em Neuropsicologia (juiz 4).

Os juízes 3 e 4 identificaram uma lacuna na versão preliminar do subteste de escrita. Estavam faltando itens muito fáceis, condizentes com a aprendizagem de crianças recém-ingressantes no 1º ano escolar. Como um dos critérios empíricos elegidos referia-se à criação de itens fáceis, os peritos criaram seis itens considerados simples, de acordo com a perspectiva teórica mencionada na etapa 1, do polo teórico (nível de aprendizagem de cada ano escolar, uso de conteúdo não regionalizado, termos não pejorativos, ortografia e grau de familiaridade dos itens).

### **Análise teórica 2**

Os seis itens criados na etapa teórica anterior foram enviados aos dois *experts* que trabalharam na criação dos itens da versão preliminar do subteste de escrita (juízes 1 e 2). Ambos deveriam analisar os seis itens, verificando a sua pertinência. Como resultado dessa análise teórica, os juízes optaram por manter um dos seis itens sugeridos, indicando outros cinco, em função de estes serem mais adequados aos critérios teóricos mencionados na etapa 1. Dessa forma, uma nova versão preliminar do subteste de escrita foi originada, a qual continha 124 itens.

### **Análise teórica 3**

Com essa segunda versão preliminar do subteste de escrita, o juiz 3 verificou novamente os itens do ponto de vista teórico. Essa nova análise teve como objetivo identificar todas as dificuldades ortográficas contidas em cada item (por exemplo, o uso de dígrafos, de *h* em início de palavra e de *e* ou *i* no fim de palavra). Essa investigação conceitual a respeito dos itens foi baseada em Moojen (2009). Assim, o juiz 3 elaborou uma ficha de avaliação, a

qual continha todas as informações referentes às dificuldades ortográficas dos itens.

### **Análise teórica 4**

A ficha de avaliação elaborada na análise teórica anterior foi enviada para outros três juízes peritos em linguagem escrita: juiz 5, com formação em Letras, mestrado e doutorado em Linguística Aplicada; juiz 6, com formação em Fonoaudiologia, especialização em Educação Infantil, mestrado e doutorado em Letras; e juiz 7, com formação em Fonoaudiologia, especialização em Fonoaudiologia Escolar, mestrado e doutorado em Letras. Esses juízes foram selecionados por apresentarem ampla experiência no construto linguagem escrita. Deveriam analisar tanto as características teóricas identificadas para cada item como a pertinência dos itens para a avaliação da habilidade de escrita de palavras. Poderiam também sugerir modificações em itens ou nas dificuldades ortográficas identificadas, bem como indicar a eliminação ou substituição de itens problemáticos.

Como resultado dessa avaliação, nenhum item foi excluído ou modificado; apenas algumas características teóricas referentes aos itens foram alteradas. A título de exemplo, observa-se o caso do item “gato”, cujas dificuldades ortográficas indicadas pelo juiz 3, quando da análise teórica 3, foram a possibilidade de substituição do grafema *g* por *gu* e a substituição do grafema *o*, em posição final, por *u*. Porém os juízes que atuaram na análise teórica 4 (juízes 5, 6 e 7) solicitaram o acréscimo da possibilidade de substituição do grafema *g* pelo grafema *h*.

Assim, a terceira versão preliminar do subteste de escrita permaneceu com 124 itens, os quais foram considerados pertinentes pelos juízes para a avaliação da habilidade de escrita de palavras.

## **Polo experimental: testagem empírica da versão preliminar do subteste de escrita do TDE II**

Com o intuito de verificar evidências de validade de construto da versão preliminar do subteste de escrita, foi realizada uma testagem empírica com estudantes do ensino fundamental de escolas públicas e privadas. O objetivo dessa testagem consistiu em verificar o comportamento empírico dos itens e sua adequabilidade para avaliar a habilidade de escrita. As etapas dessa testagem serão descritas a seguir.

### **Participantes**

Participaram do estudo 723 estudantes dos nove anos escolares do ensino fundamental de escolas públicas e privadas da região metropolitana de Porto Alegre (RS). Selecionaram-se as escolas por conveniência, sendo quatro públicas e três particulares. Foram incluídos na amostra todos os escolares com percentil igual ou maior que 5 no teste Matrizes Progressivas Coloridas de Raven: Escala Especial (Angelini, Alves, Custódio, Duarte, & Duarte, 1999) ou no teste Matrizes Progressivas de Raven: Escala Geral (Vilhena, Guntert,

& Tosi, *no prelo*). Além disso, foram excluídos sujeitos com suspeita de déficit auditivo e alterações de fala ou de linguagem. Dessa forma, deixaram a pesquisa 24 crianças por déficit intelectual, oito por conta de alterações de fala ou linguagem, quatro por serem alunos especiais da escola, dois por apresentarem diagnóstico (segundo informações da escola) de Transtorno do Espectro do Autismo, um por suspeita de comprometimento intelectual (identificada pelo questionário sociodemográfico e pelas avaliações do TDE), e um por suspeita de deficiência auditiva. Com isso, a amostra foi constituída por 684 crianças, com idades entre 6,34 e 17,75 ( $M=10,07$ ;  $DP=2,59$ ). A Tabela 1 mostra a distribuição dos estudantes por ano escolar, sexo e tipo de escola.

### Instrumentos

Os escolares foram avaliados por meio da versão preliminar do subteste de escrita e por um teste de inteligência. O teste Matrizes Progressivas Coloridas de Raven: Escala Especial (Angelini et al., 1999) foi aplicado em estudantes de até 11 anos e 9 meses de idade, e o Teste Matrizes Progressivas de Raven – Escala Geral (Vilhena et al., *no prelo*), em escolares com mais de 11 anos e 10 meses. Além disso, os responsáveis pelos estudantes responderam a um questionário acerca de dados sociodemográficos, versão adaptada do questionário disponível em Fonseca, Jacobsen, & Pureza (2015).

### Procedimentos

#### Coleta de dados

O projeto desta pesquisa foi aceito pelo Comitê de Ética da Pontifícia Universidade Católica do Rio Grande do Sul. Os estudantes foram autorizados pelos responsáveis a participar do estudo, mediante assinatura do Termo de Consentimento Livre e Esclarecido (TCLE).

No primeiro encontro foi realizada a avaliação intelectual: individual com as crianças do 1º ao 3º ano do ensino fundamental; e coletiva, em grupos de oito participantes, nos demais anos escolares.

No segundo encontro, fez-se a avaliação da escrita, por meio do ditado das palavras isoladas da versão preliminar do subteste de escrita. O examinador ditava cada palavra uma vez e repetia-a também uma vez. A criança deveria escrever as respostas em uma folha com duas colunas e linhas numeradas. Ela era orientada a não utilizar borracha, mas sim circular possíveis equívocos e escrever a forma correta ao lado. Esse procedimento teve como objetivo possibilitar as análises dos erros em estudos futuros. A avaliação ocorreu individualmente com crianças até o 3º ano escolar e coletivamente do 4º ano em diante. Nas avaliações individuais, o teste era interrompido após 10 erros consecutivos. Em relação à quantidade de palavras ditadas em cada dia de avaliação, o procedimento diferia de acordo com o ano escolar da criança, para evitar o cansaço nas crianças pequenas. Assim, em cada dia de avaliação, os estudantes do 1º ao 3º ano escolar eram expostos a itens referentes a no máximo dois anos escolares, o que correspondia a cerca de 30 itens. Ainda que a criança estivesse grafando corretamente as palavras, o teste era interrompido após a escrita dos itens referentes a dois anos escolares, e, no dia seguinte, era dada continuidade a ele. Para crianças do 4º ano escolar, o teste era aplicado em quatro dias diferentes, sendo 31 palavras ditadas em cada encontro. A partir do 5º ano, o teste era dividido em dois dias, sendo 62 itens ditados em cada momento. Posteriormente, os registros dos estudantes eram corrigidos por uma junta de três fonoaudiólogos treinados para esse fim, que deveriam classificar as respostas das crianças em “acerto” e “erro”. Ressalta-se que as avaliações aconteceram em sala silenciosa da própria escola, em horário de aula.

Tabela 1  
Caracterização sociodemográfica da amostra

Ano escolar	Escolas públicas		Escolas privadas		Faixa etária		Total
	Meninas	Meninos	Meninas	Meninos	M	DP	
1º	23	19	18	19	6,55	1,90	79
2º	23	22	39	33	7,85	0,32	117
3º	37	20	21	19	8,88	1,14	97
4º	23	32	16	16	9,87	0,57	87
5º	32	24	33	26	10,73	1,47	115
6º	30	25	15	9	11,62	2,61	79
7º	11	12	10	11	13,19	0,74	44
8º	16	19	10	8	14,28	0,95	53
9º	0	0	5	8	14,65	0,50	13
Total	195	173	167	149	10,07	2,59	684

Nota: M=media; DP=desvio-padrão.

## Análise de dados

Para análise dos dados, os itens do teste foram tratados como variáveis dicotômicas de acerto e erro. Para a operacionalização do estudo, realizaram-se procedimentos analíticos, como a análise da dimensionalidade do instrumento, a análise da fidedignidade e a análise empírica dos itens, com base nos critérios sugeridos por Pasquali (1999).

Os procedimentos de análise dos dados contaram com três etapas. Inicialmente, fez-se a avaliação da dimensionalidade do subteste, por meio de análises fatoriais exploratórias dos eixos principais das correlações tetracóricas dos itens (Revelle, 2015). A segunda etapa constituiu a análise da consistência interna do instrumento, por meio do coeficiente alfa, o qual foi utilizado com o intuito de verificar a precisão do instrumento (Zanon & Hauck Filho, 2015). Por fim, a última fase da análise de dados consistiu na TRI, que foi utilizada para estimar o nível de habilidade dos sujeitos e os parâmetros de discriminação e dificuldade dos itens. Neste trabalho, os pressupostos de monotonicidade (a probabilidade de acertar determinado item cresce proporcionalmente ao aumento da habilidade do participante), independência local (os itens do teste não influenciam um no outro para a obtenção da resposta do sujeito) e unidimensionalidade (o conjunto de itens diz respeito a um único traço latente) foram investigados, a fim de proceder-se às análises de TRI paramétricas (DeMars, 2010; Finch Jr. & French, 2015; van der Linden & Hambleton, 1997).

O subteste de escrita foi dividido em duas versões, de acordo com os pressupostos teóricos referentes à distribuição dos itens por anos escolares do ensino fundamental. Dessa forma, julgou-se oportuno direcionar os itens referentes ao 1º, 2º, 3º e 4º anos escolares para a versão A do subteste (65 itens) e os itens pertencentes ao 5º, 6º, 7º, 8º e 9º anos para a versão B (59 itens). Com essa divisão dos itens, foi possível constatar a unidimensionalidade das versões A e B, o que é condição necessária para o uso da TRI. Portanto, as análises subsequentes tiveram como objetivo identificar as características psicométricas de cada uma das versões, o que será descrito a seguir.

Para a realização das análises, utilizaram-se os *softwares* Statistical Package for the Social Sciences (SPSS) e o *software* R (R Core Team, 2015), com as funções dos pacotes *psych* (Revelle, 2015), *ltm* (Rizopoulos, 2006) e *mokken* (van der Andries, 2007).

## Resultados

### Polo analítico: versão A do subteste de escrita do TDE II — dimensionalidade, fidedignidade, TRI

Inicialmente, a versão A do subteste de escrita contou com 65 itens. O estudo da dimensionalidade do subteste foi feito por meio da análise fatorial exploratória (AFE)

da matriz de correlações tetracóricas dos itens, método adequado para itens dicotômicos. Foi utilizado o método de extração de eixos principais (*principal axis factoring*). Os pressupostos para a realização da AFE foram alcançados, uma vez que o teste de Kaiser Meyer Olkin (KMO) foi de 0,99, e o teste de esfericidade de Bartlett foi significativo,  $\chi^2(2,08)=37.620,89, p<0,0001$ , permitindo confirmar a adequação da análise fatorial para o tratamento dos dados. Para avaliar a pertinência de cada item com a dimensão avaliada, foram considerados itens que: possuíam o valor absoluto de carga fatorial principal acima de 0,30 (para amostras acima de 350 participantes e com significância baseada em  $p<0,05$  e poder estatístico de 80%); apresentaram cargas fatoriais similares em dois ou mais fatores em um mesmo item, sendo a diferença entre valores absolutos das cargas fatoriais dos itens maior do que 0,10; e tinham relação inequívoca com sua dimensão, ou seja, demonstravam convergência teórica com os demais itens do mesmo fator (Hair Jr., Black, Babin, & Anderson, 2010; Laros, 2005).

A análise fatorial apontou um fator explicando 71% da variância dos dados, e as cargas fatoriais dos itens ficaram entre 0,43 e 0,95. Nessa etapa, nenhum item foi eliminado, uma vez que todas as cargas fatoriais obtiveram valores acima de 0,40. O alfa obtido foi de 0,98, o que demonstra uma propriedade de fidedignidade do subteste, já que, segundo Zanon e Hauck Filho (2015), valores acima de 0,90 são considerados excelentes.

A seguir, a análise empírica dos itens aconteceu por meio da TRI. A investigação dos seus pressupostos foi testada mediante a análise da escalabilidade do teste, seguindo o procedimento proposto por *mokken* (Rusch, Mair, & Hatzinger, 2013; van der Andries, 2007; van der Linden & Hambleton, 1997). Nessa abordagem, valores de escalabilidade acima de 0,30 para o conjunto de itens são considerados aceitáveis. A versão A do subteste de escrita obteve o valor de 0,68, sendo, portanto, satisfatório. A monotonicidade dos itens foi avaliada, e todos os itens obtiveram valores acima de 0,46, o que é tido como adequado para a utilização de modelos paramétricos da TRI (Junker & Sijtsma, 2000). Dessa forma, o conjunto de itens da versão A é satisfatório para a realização das análises via TRI.

Quatro modelos de TRI paramétricos foram testados no tratamento dos dados: Rash, modelo de um parâmetro (1PL), modelo de dois parâmetros (2PL) e modelo de três parâmetros (3PL). Esses quatro modelos foram avaliados por intermédio do teste de hipótese estatística na comparação de valores de verossimilhança, do critério de informação de Akaike (AIC) e do critério de informação bayesiano (BIC). A comparação entre o modelo 2PL (verossimilhança igual a -14.622,90, AIC igual a 29.505,81 e BIC igual a 30.091,36) foi significativamente melhor que o modelo Rasch (verossimilhança igual a -15.935,80, AIC igual a 32.001,61 e BIC igual a 32.294,39) e que o

modelo 1PL (verossimilhança igual a -15.044,84, AIC igual a 30.221,68 e BIC igual a 30.518,96). Ao comparar-se os modelos 2PL e 3PL, optou-se pelo primeiro, em virtude de o modelo 3PL ter apresentado solução instável, já que a matriz hessiana convergiu em valores infinitos, produzindo *missing values*.

Posteriormente, procedeu-se à análise de resíduos do subteste, uma vez que a TRI requer uma investigação formal sobre o funcionamento dos itens enquanto medidas de traço latente (Glas, 2016; Reise, 1990; Wells & Hambleton, 2016). A função *item.fit* (Rizopoulos, 2006) foi usada para comparar os valores previstos de cada item do modelo 2PL com os valores empíricos coletados por meio do teste de  $\chi^2$ . Nessa etapa, 15 itens apresentaram desajuste e foram excluídos. Assim, a versão A final do subteste de escrita resultou em 50 itens.

Identificou-se, porém, a necessidade de diminuir a quantidade de itens, considerando que o TDE se destina a ser um instrumento de avaliação breve do desempenho escolar. Para tanto, realizou-se nova revisão dos itens, a fim de eliminar os itens menos relevantes para a avaliação do construto. A eliminação desses itens aconteceu seguindo critérios empíricos e teóricos. Primeiramente, distinguiram-se os itens que mensuravam o mesmo nível de habilidade (itens com

níveis de dificuldade muito próximos). Depois, esses itens redundantes passaram por nova análise teórica, a qual foi desenvolvida por *experts* em linguagem escrita. O juiz deveria analisar os pares de itens com dificuldades semelhantes, optando pela permanência de um em detrimento do outro. Para isso, levaram-se em conta as características teóricas dos itens, bem como os valores de discriminação.

Diante dessa análise, a versão A final do subteste de escrita contou com 40 itens. O modelo 2PL final obteve valor de verossimilhança igual a -9.676,41, valor de AIC igual a 19.512,83 e valor de BIC de 19.873,17, e a dificuldade dos itens variou de -1,4146 a 2,1218, conforme Tabela 2.

Os escores dos alunos de 1º a 4º ano ficaram entre -2,14 e 1,94 na versão A, sendo a média -0,05 ( $DP=0,98$ ). Nesse intervalo de habilidade, a versão A do subteste de escrita possui 95,5% de informação sob a curva de informação do teste (Figura 1).

#### **Polo analítico: versão B do subteste de escrita do TDE II — dimensionalidade, fidedignidade, TRI**

A versão B do subteste de escrita foi composta inicialmente por 59 itens. A análise fatorial exploratória apontou uma dimensão explicando 56% de variância

Tabela 2  
Dificuldade e discriminação dos itens da versão A do subteste de escrita

Item	Dificuldade	Discriminação	Item	Dificuldade	Discriminação
E1	-1,415	2,054	E43	-0,098	4,218
E5	-1,412	2,066	E32	-0,082	3,208
E2	-1,341	2,467	E41	-0,072	4,248
E3	-1,335	2,773	E25	-0,046	2,916
E9	-1,120	2,045	E28	0,047	3,065
E17	-0,704	3,281	E27	0,075	2,542
E31	-0,611	2,526	E60	0,086	3,203
E20	-0,565	2,796	E44	0,101	2,186
E21	-0,499	3,050	E35	0,168	2,711
E11	-0,414	1,538	E65	0,312	0,993
E22	-0,396	3,085	E59	0,333	1,466
E46	-0,389	2,543	E33	0,374	2,625
E24	-0,344	2,327	E54	0,501	3,610
E23	-0,326	2,591	E55	0,535	3,404
E49	-0,321	3,949	E51	0,674	1,582
E39	-0,222	3,456	E56	0,780	3,619
E34	-0,174	2,470	E30	0,972	1,999
E50	-0,144	2,869	E38	1,117	2,488
E12	-0,116	2,370	E57	2,119	1,687
E42	-0,101	3,834	E45	2,122	0,999

dos dados ( $KMO=0,94$ , teste de esfericidade de Bartlett  $\chi^2(1,71)=19.327,03$ ,  $p<0,0001$ ). Nessa etapa um item foi excluído, por possuir carga fatorial abaixo de 0,30. Já a carga fatorial dos demais itens variaram de 0,44 a 0,86. O alfa obtido foi de 0,86, endossando a consistência interna do subteste. O pressuposto da escalabilidade ( $H=0,52$ ) e da monotonicidade dos itens (todos com  $H_i>0,39$ ) foi alcançado, permitindo a utilização de modelos de TRI para a análise empírica dos itens da versão B do subteste de escrita (Junker & Sijtsma, 2000; Rusch et al., 2013; van der Linden & Hambleton, 1997).

Seguindo os mesmos procedimentos analíticos descritos para a comparação de modelos de TRI da versão A, o modelo 2PL (verossimilhança igual a -8.431,62, AIC igual a 17.095,24 e BIC igual a 17.617,74) mostrou-se estatisticamente mais satisfatório para o tratamento dos dados, em comparação aos modelos Rasch (verossimilhança igual a -8.932,41, AIC igual a 17.980,81 e BIC igual a 18.242,06) e 1PL (verossimilhança igual a -8.454,35, AIC igual a 17.126,69 e BIC igual a 17.292,44). Ao comparar os modelos 2PL e 3PL, optou-se pelo primeiro, em virtude de o modelo 3PL ter apresentado novamente solução instável, já que a matriz hessiana convergiu em valores infinitos, produzindo *missing values*.

Pode-se inferir que a produção de solução instável pelo modelo 3PL nas duas versões do subteste de escrita se deve ao tipo de resposta atribuída pelo aluno, em que

não há espaço para acerto ao acaso (“chute”). Como o modelo de três parâmetros visa diminuir o efeito do chute e no subteste de escrita não existe essa possibilidade, é comum o surgimento dessa solução instável (Ayala, 2009; DeMars, 2010). Portanto, procedeu-se às análises utilizando o modelo 2PL.

A investigação formal da adequação dos itens ao modelo foi realizada mediante novamente a função *item fit*. Como resultado dessa última análise, nove itens foram excluídos, e a versão B do subteste de escrita contabilizou 50 itens.

Porém, diante do número elevado de itens, o mesmo procedimento aplicado na exclusão de itens da versão A foi usado na versão B. Assim, por meio de critérios empíricos e teóricos, 10 itens foram excluídos. Com isso, o modelo 2PL final contou com 40 itens (Tabela 3, cujas dificuldades variaram de 0,41 a 2,98, com valor de verossimilhança igual a -5.779,61, AIC igual a 11.719,22 e BIC igual a 12.079,57).

Na versão B do subteste, os escores dos alunos de 5º a 9º ano variaram de -0,76 a 2,47 ( $M=0,66$ ,  $DP=0,76$ ). Nesse intervalo de habilidade, a versão B do subteste de escrita possui 79,67% de informação sob a curva do teste (Figura 2). Vale ressaltar que a versão B do subteste de escrita alcança escores nos valores de -0,76 a 3,51, portanto os alunos de 5º a 9º ano não conseguiram alcançar o escore máximo do subteste.

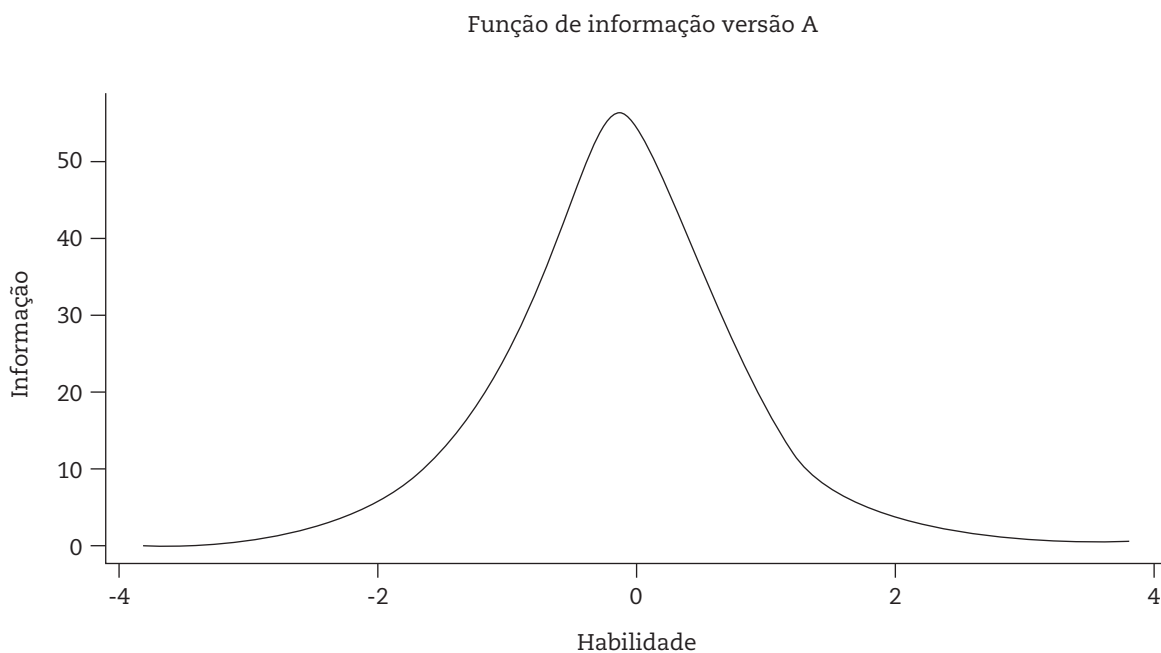


Figura 1. Curva de informação da versão A do Teste de Desempenho Escolar (TDE II)



Tabela 3  
Dificuldade e discriminação dos itens da versão B do subteste de escrita

Item	Dificuldade	Discriminação	Item	Dificuldade	Discriminação
E100	0,415	2,770	E118	1,555	2,451
E83	0,549	2,711	E90	1,624	2,234
E66	0,563	3,128	E93	1,733	1,670
E85	0,578	3,135	E87	1,750	3,505
E70	0,655	2,753	E102	1,820	3,378
E79	0,779	3,427	E103	1,865	3,881
E72	0,902	2,337	E117	2,224	3,157
E73	0,967	2,808	E104	2,233	4,134
E71	0,976	1,279	E86	2,257	3,300
E68	0,987	2,623	E121	2,303	3,253
E92	1,003	2,826	E123	2,359	2,060
E91	1,033	3,149	E113	2,405	2,023
E75	1,133	2,356	E96	2,438	2,228
E89	1,169	3,357	E95	2,617	2,650
E78	1,208	2,427	E94	2,759	2,239
E119	1,237	2,961	E120	2,760	2,072
E69	1,342	3,392	E116	2,828	1,913
E106	1,361	3,056	E114	2,853	3,564
E112	1,429	3,008	E110	2,878	0,975
E107	1,535	2,362	E115	2,986	2,318

Função de informação versão B

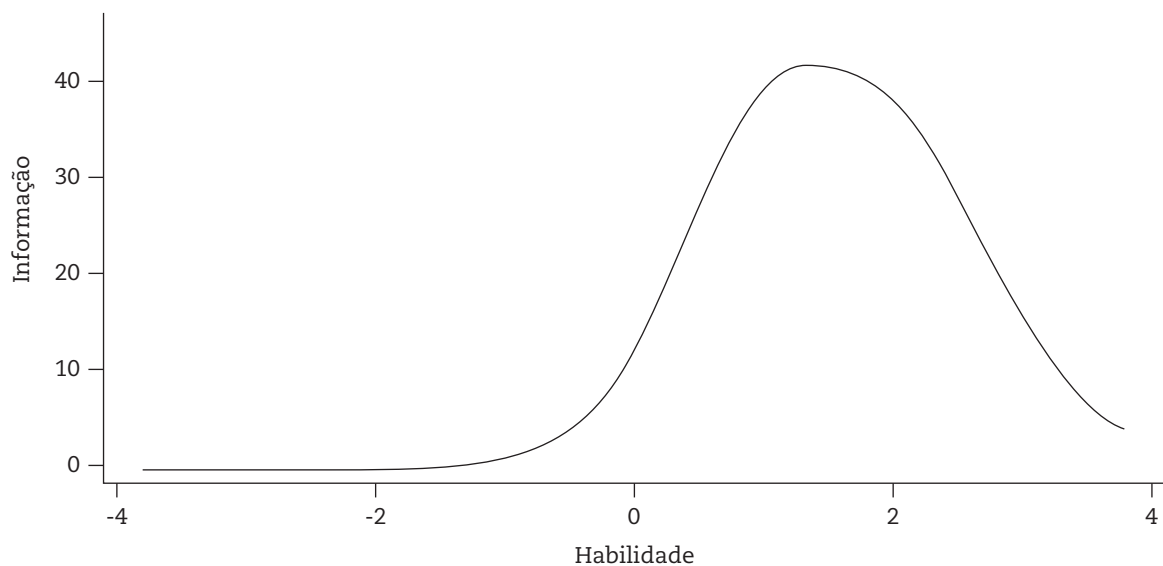


Figura 2. Curva de informação da versão B do Teste de Desempenho Escolar (TDE II)

## Discussão

O presente artigo apresentou o processo de desenvolvimento do subteste de escrita do TDE II. Sua elaboração ocorreu mediante etapas, as quais são sugeridas por Pasquali et al. (2010) para testes referentes a construto. As etapas mencionadas fundamentam-se em três grandes polos — teórico, experimental e analítico —, que foram conduzidos com base em critérios rigorosos, conferindo confiança no processo de desenvolvimento do subteste.

O subteste de escrita do TDE II demonstra avaliar adequadamente crianças dos nove anos escolares do ensino fundamental, apresentando boa discriminação entre os alunos das amostras A e B. No subteste original, pesquisas anteriores (Brito et al., 2012; Knijnik et al., 2014; Giacomoni et al., 2015) observaram que o instrumento não estava discriminando suficientemente alunos com níveis de habilidade baixo e alto. Diferentemente pôde ser constatado no subteste de escrita do TDE II, uma vez que a curva de informação das versões A e B indica estar avaliando apropriadamente sujeitos de todo o *continuum* da habilidade escrita de palavras. Além disso, estudantes dos últimos anos escolares não atingiram a pontuação máxima da versão B. Segundo Urbina (2007), esse dado supõe que a dificuldade do subteste está satisfatória.

Provavelmente esse refinamento verificado no novo subteste se deve ao fato de suas etapas de desenvolvimento terem sido conduzidas com rigor metodológico. No que tange ao polo teórico, destaca-se a importância da análise de juízes *experts*, a qual foi fundamental para criar e selecionar os itens, bem como para levar em conta a pertinência deles para a avaliação do construto escrita de palavras. Essa análise de juízes, segundo Pasquali et al. (2010), confere validade de conteúdo ao instrumento.

Porém essa pertinência dos itens foi constatada apenas em nível teórico na etapa supracitada, uma vez que os juízes não eram representativos da população a que o instrumento se destinava (Pasquali et al., 2010). Portanto, as evidências de validade empírica deram-se no polo analítico, no qual os dados coletados no polo empírico foram analisados. Nesse ponto, constatou-se que, pelo valor do alfa encontrado, ambas as versões do subteste de escrita do TDE II apresentaram evidências de fidedignidade, indicando consistência e precisão dos resultados (Kaplan & Saccuzzo, 2013; Urbina, 2007).

Ainda em relação ao polo analítico, destacam-se dois pontos. O primeiro é o uso da psicometria clássica (teoria clássica dos testes – TCT) e moderna (TRI) na análise dos dados, o que aprimorou o presente trabalho.

No tocante à TCT, a análise fatorial, por meio da análise das intercorrelações dos itens, permite que um grande número de variáveis possa ser explicado por um número menor de variáveis hipotéticas, não observadas (Field, Miles, & Field, 2012; Laros, 2005). Muitos pesquisadores das ciências sociais reconhecem o relacionamento próximo entre análise fatorial e questões de validade interna do instrumento ou de validade de construto (Hutz, Bandeira, & Trentini, 2015; Izquierdo, Olea, & Abad, 2014), porém a TRI tem se mostrado mais eficaz para a análise empírica dos itens, pois níveis de dificuldade e discriminação não dependem da amostra quando os seus pressupostos são alcançados (DeMars, 2010; Pasquali et al., 2010). O segundo ponto que merece destaque no polo analítico se refere à realização de nova análise teórica dos itens após as análises psicométricas. Essa última análise teórica teve como objetivo diminuir a quantidade de itens de cada versão do subteste, a fim de que com ela fosse possível realizar a avaliação da habilidade de escrita de palavras de forma breve.

No que tange ao polo experimental, a testagem dos itens em uma amostra representativa da população a que o teste se destina contribuiu para a obtenção das evidências de validade de critério. Graças a essa testagem, foi possível realizar mudanças finais na ordem de apresentação dos itens, para que estes seguissem uma ordem crescente de dificuldade. Giacomoni et al. (2015) perceberam que os itens do subteste de escrita original não estavam dispostos em uma ordem adequada, reforçando a importância desse cuidado quando do desenvolvimento de sua versão atualizada.

Por fim, após as análises conduzidas, os resultados sugerem que o instrumento é relevante para a prática clínica, educacional e científica, uma vez que, segundo Knijnik et al. (2013), há a carência de instrumentos científicos que avaliam o desempenho escolar. Ressalta-se que o TDE II passou a ser considerado um instrumento de avaliação breve e não mais um *screening*, como a versão original. Por meio do subteste de escrita do TDE II, é possível identificar as crianças que apresentam dificuldade na habilidade de escrita de palavras, bem como quais dificuldades de conversão fonema-grafema e/ou ortográficas possuem. Assim, o referido subteste pode funcionar como uma triagem escolar, mas também como o primeiro procedimento da avaliação clínica do desempenho em escrita.

Para estudos futuros, recomendam-se pesquisas que busquem outras evidências de validade no subteste de escrita do TDE II, como as validades convergente e teste-reteste, além de estudos sobre dados normativos por idade e tipo de escola.

## Referências

- Abbott, R., Berninger, V. W., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology, 102*(2), 281-298. doi: 10.1037/a0019318
- Angelini, A. L., Alves, I. C. B., Custódio, E. M., Duarte, W. F., & Duarte, J. L. M. (1999). *Manual matrizes progressivas coloridas de Raven: escala especial*. São Paulo: Centro Editor de Testes e Pesquisas em Psicologia.
- Ayala, R. J. (2009). *The theory and practice of Item Response Theory*. Nova York: The Guilford Press.
- Batista, A. O., & Capellini, S. A. (2011). Desempenho ortográfico de escolares do 2º ao 5º ano do ensino privado do município de Londrina. *Psicologia e Argumento, 29*(67), 411-425.
- Brito, L. O., Ambiel, R. A. M., Pacanaro, S. V., Grisard, E., Alves, G. A. S., Rabelo, I. S., Leme, I. F. A. S. (2012). Relação das variáveis idade e escolaridade com desempenho escolar de estudantes de ensino fundamental. *Avaliação Psicológica, 11*(1), 83-93.
- Capellini, S. A., Oliveira, A. M., & Cuetos, F. (2010). *Prolec – Provas de Avaliação dos Processos de Leitura*. Itatiba: Casa do Psicólogo.
- Capovilla, A., Smythe, I., Capovilla, F., & Everatt, J. (2001). Adaptação brasileira do International Dyslexia Test: perfil cognitivo de crianças com escrita pobre. *Temas sobre Desenvolvimento, 10*(57), 30-37.
- Capovilla, A. G. S., & Capovilla, F. C. (2003). *Problemas de leitura e escrita: como identificar, prevenir e remediar numa abordagem fônica*. 3.ª edição. São Paulo: Memnon.
- Conrad, N. J., Harris, N., & Williams, J. (2013). Individual differences in children's literacy development: The contribution of orthographic knowledge. *Reading and Writing, 26*, 1223-1239. doi: 10.1007/s11145-012-9415-2
- Conselho Federal de Psicologia (2003). *Resolução CFP n.º 002/2003*. Recuperado em 2 de dezembro de 2015, de [http://site.cfp.org.br/wp-content/uploads/2003/03/resolucao2003\\_02\\_Anexo.pdf](http://site.cfp.org.br/wp-content/uploads/2003/03/resolucao2003_02_Anexo.pdf).
- Costa, L. J. C., Edwards, C. N., & Hooper, S. R. (2016). Writing disabilities and reading disabilities in elementary school students. *Learning Disability Quarterly, 39*(1), 17-30. doi: 10.1177/0731948714565461
- Crossley, S. A., Weston, J. L., Sullivan, S. T. M., & McNamara, D. S. (2011). The development of writing proficiency as a function of grade level: A linguistic analysis. *Written Communication, 28*(3) 282-311. doi: 10.1177/0741088311410188
- Cunha, V. L. O., & Capellini, S. A. (2009). *Prohmele – Provas de Habilidades Metalinguísticas e de Leitura*. Rio de Janeiro: Revinter.
- Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: Norms for name, agreement, familiarity and visual complexity. *Journal of Experimental Child Psychology, 65*(2), 171-237. doi: 10.1006/jecp.1996.2356
- DeMars, C. (2010). *Item Response Theory: Understanding statistics measurement*. Nova York: Oxford University Press.
- Ferreira, A. T. B., & Leal, T. F. (2007). Avaliação na escola e ensino da língua portuguesa: introdução ao tema. Em B. Marcurschi & L. Suassuna (Eds.). *Avaliação em língua portuguesa: contribuições para a prática pedagógica*. (pp. 11-26). Belo Horizonte: Autêntica.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Londres: Sage Publications.
- Finch Jr., W. H., & French, B. F. (2015). *Latent variable modeling with R*. Nova York: Routledge.
- Fonseca, R. P., Jacobsen, G. M., & Pureza, J. R. (2016). O que um bom teste neuropsicológico deve ter? Em J. F. Salles, V. G. Haase & L. F. Malloy-Diniz. *Neuropsicologia do desenvolvimento: infância e adolescência*. Porto Alegre: Artmed.
- Fonseca, R. P., Salles, J. F., & Parente, M. A. M. P. (2009). *Instrumento de Avaliação Neuropsicológica Breve – NEUPSILIN*. São Paulo: Vetor.
- Giacomoni, C. H., Athayde, M. L., Zanon, C., & Stein, L. M. (2015). Teste de Desempenho Escolar: evidências de validade do subteste de escrita. *Psico-USF, 20*(1), 133-140. doi: 10.1590/1413-82712015200112
- Glas, C. A. W. (2016). Frequentist Model-Fit Tests. Em W. J. van der Linden (Ed.). *Handbook of Item Response Theory: Statistical tools*. v. 2. (pp. 313-361). Boca Raton: CRC Press.
- Hair Jr., J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. New York, NY: Pearson Prentice Hall.
- Hutz, C. S., Bandeira, D. R., & Trentini, C. M. (2015). *Psicometria*. Porto Alegre: Artmed.
- International Test Commission. (2000). *Diretrizes para o uso de testes* (Instituto Brasileiro de Avaliação Psicológica, Trad., 2003). Recuperado em 15 de setembro de 2014, de [https://www.intestcom.org/files/guideline\\_test\\_use\\_portuguese\\_brazil.pdf](https://www.intestcom.org/files/guideline_test_use_portuguese_brazil.pdf)
- Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analysis in validation studies: Uses and recommendations. *Psicothema, 26*(3), 395-400. doi: 10.7334/psicothema2013.349
- Junker, B. W., & Sijtsma, K. (2000). Latent and manifest monotonicity in Item Response Models. *Applied Psychological Measurement, 24*(1), 65-81. doi: 10.1177/01466216000241004
- Kaplan, R. M., & Saccuzzo, D. P. (2013). *Psychological testing: Principles, applications, and issues*. 8.ª edição. Belmont: Cengage Learning.
- Knijnik, L. F., Giacomoni, C., & Stein, L. M. (2013). Teste de Desempenho Escolar: um estudo de levantamento. *Psico-USF, 18*(3), 407-416. doi: 10.1590/S1413-82712013000300007
- Knijnik, L. F., Zanon, C., Giacomoni, C. H., & Stein, L. M. (2014). Avaliação dos subtestes de leitura e escrita do Teste de Desempenho Escolar através da Teoria de Resposta ao Item. *Psicologia: Reflexão e Crítica, 27*(3), 481-490. doi: 10.1590/1678-7153.201427308
- Laros, J. A. (2005). O uso da análise fatorial: algumas diretrizes para pesquisadores. Em L. Pasquali (Ed.). *Análise fatorial para pesquisadores*. (pp. 163-184). Brasília: LabPAM.
- Loveall, S. J., Channell, M. M., Phillips, B. A., & Connors, F. A. (2013). Phonological recoding, rapid automatized naming, and orthographic knowledge. *Journal of Experimental Child Psychology, 116*(3), 738-746. doi: 10.1016/j.jecp.2013.05.009
- Moojen, S. M. P. (2009). Linguagem escrita. Em S. M. P. Moojen. *A escrita ortográfica na escola e na clínica: teoria, avaliação e tratamento*. São Paulo: Casa do Psicólogo.
- Pasquali, L. (1999). Testes referentes a construto: teoria e modelo de construção. Em L. Pasquali (Ed.). *Instrumentos psicológicos: manual prático de elaboração*. (pp. 37-73). Brasília: LabPAM.
- Pasquali, L. (2010). *Instrumentação psicológica: fundamentos e práticas*. Porto Alegre: Artmed.
- Pinheiro, A. M. V. (1995). Reading and spelling development in Brazilian Portuguese. *Reading & Writing*, special issue on Literacy Acquisition, 7, 111-138.
- Pinheiro, A. M. V., Rothe-Neves, R. (2001). Avaliação cognitiva de leitura e escrita: as tarefas de leitura em voz alta e ditado. *Psicologia: Reflexão e Crítica, 14*(2), 399-408.
- R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Viena, Áustria. Recuperado de <http://www.R-project.org/>

- Reise, S. P. (1990). A comparison of item- and person-fit methods of assessing model-data fit in IRT. *Applied Psychological Measurement*, 14(2), 127-137. doi: 10.1177/014662169001400202
- Revelle, W. (2015). Psych: Procedures for Personality and Psychological Research. Evanston, Illinois. Recuperado de <http://cran.r-project.org/package=psych>
- Rizopoulos, D. (2006). ltm: An R package for latent variable modeling and Item Response Theory analyses. *Journal of Statistical Software*, 17(5), 1-25. Recuperado de <http://www.jstatsoft.org/v17/i05/>
- Rueda, F. J. M., Noronha, A. P. P., Sisto, F. F., Santos, A. A. A., & Castro, N. R. (2012). *Escala de inteligência Wechsler para crianças – WTSC IV*. São Paulo: Casa do Psicólogo.
- Rusch, T., Mair, P., & Hatzinger, R. (2013). Psychometrics with R: A review Of CRAN packages for Item Response Theory. *Discussion Paper Series of the Center for Empirical Research Methods*, 2.
- Salles, J. F. (2005). *Habilidades e dificuldades de leitura e escrita em crianças de 2.ª série: abordagem neuropsicológica cognitiva* (Tese de doutorado não publicada, Pós-Graduação em Psicologia do Desenvolvimento, Universidade Federal do Rio Grande do Sul, Porto Alegre, R.S.) Recuperado em 15 de setembro de 2014, de <http://www.lume.ufrgs.br/bitstream/handle/10183/4197/000453537.pdf?sequence=1>
- Salles, J. F., & Parente, M. A. M. P. (2007). Avaliação da leitura e escrita de palavras em crianças de 2.ª série: abordagem neuropsicológica cognitiva. *Psicologia, Reflexão e Crítica*, 20, 218-226. doi: 10.1590/S0102-79722007000200007
- Stein, L. M., Giacomoni, C. H., & Fonseca, R. P. (2015). *TDE II – Teste de Desempenho Escolar II. Relatório técnico não publicado*.
- Stein, L. M. (1994). *TDE – Teste de Desempenho Escolar: manual para aplicação e interpretação*. São Paulo: Casa do Psicólogo.
- Urbina, S. (2007). *Fundamentos da testagem psicológica*. Porto Alegre: Artmed.
- van der Andries, A. L. (2007). Mokken Scale Analysis in R. *Journal of Statistical Computing*, 20(11).
- van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern Item Response Theory*. Nova York: Springer.
- Vilhena, C., Guntert, I. B., & Tosi, A. D. (no prelo). *Teste Matrizes Progressivas de Raven*. São Paulo: Casa do Psicólogo.
- Wells, C. S., & Hambleton, R. K. (2016). Model Fit with Residual Analyses. Em W. J. van der Linden (Ed.). *Handbook of Item Response Theory: statistical tools*. v. 2. (pp. 395-413). Boca Raton: CRC Press.
- Zanon, C., & Hauck Filho, N. (2015). Fidedignidade. Em C. S. Hutz, D. R. Bandeira & C. M. Trentini (Eds.). *Psicometria*. (pp. 85-97). Porto Alegre: Artmed.

recebido em dezembro de 2015  
reformulado em maio de 2016  
aprovado em junho de 2016

## Sobre os autores

**Marcia de Lima Athayde** é graduada em Fonoaudiologia pela Universidade Federal de Santa Maria e Fonoaudióloga da Universidade Federal do Rio Grande do Sul.

**Claudia Hofheinz Giacomoni** é Graduada em Psicologia pela Pontifícia Universidade Católica do Rio Grande do Sul e Professora Adjunta Departamento de Psicologia do Desenvolvimento e da Personalidade e do Programa de Pós-graduação em Psicologia da Universidade Federal do Rio Grande do Sul.

**Rochele Paz Fonseca** é graduada em Fonoaudiologia pela Universidade Luterana do Brasil e Professora Adjunta da Faculdade de Psicologia e do Programa de Pós-Graduação em Psicologia da Pontifícia Universidade Católica do Rio Grande do Sul.

**Euclides José de Mendonça Filho** é graduado em Psicologia pela Universidade Federal da Bahia e bolsista CNPQ de mestrado em psicologia pela Universidade Federal do Rio Grande do Sul.

**Lilian Milnitsky Stein** é graduada em Psicologia pela Universidade Federal do Rio Grande do Sul e Professora Adjunta da Faculdade de Psicologia e do Programa de Pós-Graduação em Psicologia da Pontifícia Universidade Católica do Rio Grande do Sul.