



Teste Informatizado e Dinâmico de Escrita: Evidências de Validade Baseadas na TRI

Léia Gonçalves Gurgel¹

Universidade Federal de Santa Catarina, Florianópolis-SC, Brasil

Livia Padilha de Teixeira

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

Vanessa Kaiser , Monica Maria Celestina de Oliveira , Caroline Tozzi Reppold

Universidade Federal de Ciências da Saúde de Porto Alegre-RS, Brasil

RESUMO

Este estudo objetiva buscar evidências de precisão e de validade com base na estrutura interna do Teste Informatizado e Dinâmico de Escrita, utilizando a Teoria de Resposta ao Item. Foram incluídos 304 participantes de Porto Alegre, entre 10 e 17 anos. Utilizou-se o TIDE, cujo objetivo é avaliar o potencial de aprendizagem em escrita narrativa de adolescentes. Verificou-se a dimensionalidade dos módulos em análise por meio de análise fatorial, satisfazendo critério para a realização das análises por meio da TRI. O módulo pré-teste foi ajustado para o modelo Logístico de dois parâmetros, com alfa de Cronbach de 0,82. Todos os itens apresentaram valores adequados de discriminação, entre 1,88 e 3,49. A dificuldade dos itens variou de -0,01 a 1,27. O módulo instrucional foi ajustado pelo modelo de resposta gradual de Samejima, apresentando alfa de Cronbach foi de 0,91. Observou-se, portanto, evidências de validade com base na estrutura interna do TIDE.

Palavras-chave: estudos de validade; psicometria; adolescentes; teoria de resposta ao item.

ABSTRACT – Computerized and Dynamic Writing Test: Evidence of validity based on IRT

This study aimed to investigate evidence of accuracy and validity based on the internal structure of the Computerized and Dynamic Writing Test, using the Item Response Theory. A total of 304 participants from Porto Alegre, aged between 10 and 17 years, were included. The aim of this study was to evaluate the adolescents' potential for learning in narrative writing. The dimensionality of the modules under analysis was verified by means of factorial analysis, satisfying criterion for the accomplishment of the analyses through IRT. The pre-test module was adjusted for the two-parameter logistic model, with Cronbach's Alpha of .82. All items presented adequate discrimination values between 1.88 and 3.49. The difficulty of the items ranged from -0.01 to 1.27. The instructional module was adjusted using Samejima's graded response model, presenting a Cronbach's alpha of .91. Accordingly, evidence of validity based on the internal structure of the CDWT was observed.

Keywords: validity studies; psychometry; adolescents; item response theory.

RESUMEN – Test Informatizado y Dinámico de Escritura: Evidencias de validez con base en la TRI

Este estudio objetiva buscar evidencias de precisión y de validez con base en la estructura interna del Test Informatizado y Dinámico de Escritura, utilizando la Teoría de Respuesta al Ítem. Participaron 304 participantes de Porto Alegre, entre 10 y 17 años. El TIDE fue utilizado con el objetivo de evaluar el potencial de aprendizaje en la escritura narrativa de adolescentes. La dimensionalidad de los módulos en análisis se averiguó por medio de análisis factorial, satisfaciendo los criterios para la realización de los análisis por medio de la TRI. El módulo pre-test se ajustó al modelo Logístico de dos parámetros, con *Alpha de Cronbach* de 0,82. Todos los ítems presentaron valores adecuados de discriminación, entre 1,88 y 3,49. La dificultad de los ítems sufrió variación de -0,01 a 1,27. El módulo instrucional fue ajustado por el modelo de respuesta gradual de Samejima, presentando Alpha de Cronbach de 0,91. Por lo tanto, se observaron evidencias de validez con base en la estructura interna del TIDE.

Palabras clave: estudios de validez; psicometría; adolescentes; teoría de respuesta al ítem.

A avaliação dinâmica é uma abordagem que promove a avaliação do potencial de aprendizagem dos sujeitos, unindo ensino e avaliação, por meio da mediação de um avaliador ao longo da realização do teste. É baseada na teoria sociocultural de Vygotsky e pressupõe que um mediador ofereça suporte intencional quando surgirem

dificuldades ao longo da realização da tarefa (Poehner et al., 2015). Nessa modalidade, diferentemente dos testes convencionais, há integração entre o desempenho e a mediação, favorecendo a interpretação das capacidades dos respondentes. Sua origem deu-se com os trabalhos de Reuven Feuerstein, na década de 1970, com o objetivo

¹ Endereço para correspondência: Laboratório de Pesquisa em Avaliação Psicológica. Universidade Federal de Ciências da Saúde de Porto Alegre. Rua Sarmento Leite, 245, Prédio II, Sala 116 e 117, Bairro Centro Histórico, 90050-170, Porto Alegre, RS. E-mail: leiagg@gmail.com
Artigo derivado da Tese de doutorado de Léia Gonçalves Gurgel com orientação de Caroline Tozzi Reppold, defendida em 2017 no programa de pós-graduação 'em Ciências da Saúde' da 'Universidade Federal de Ciências da Saúde de Porto Alegre'.

de que pudesse ser aplicada por educadores e profissionais de saúde, de forma a considerar explicações alternativas para um possível desempenho fraco em aprendizagem, por exemplo (Enumo, 2005). Em texto clássico, Tzurriel (2000) aponta a avaliação dinâmica como voltada à avaliação do pensamento, da percepção e da aprendizagem, diferindo-se das testagens convencionais por possibilitar mediação ou intervenção associada à avaliação. As estratégias de mediação fornecem melhor identificação da modificabilidade cognitiva se comparadas aos testes estáticos (Ferriolli et al., 2001).

Em âmbito nacional, as pesquisas com esse tipo de avaliação ainda estão em estágios iniciais. No entanto, instrumentos baseados nessa perspectiva já se mostram efetivos na ligação entre aprendizagem, ensino e avaliação, possibilitando que os sujeitos possam ser avaliados com vistas à intervenção, ao passo que a avaliação assistida tem como pressuposto teórico a própria mediação (Alavi & Taghizadeh, 2014). Alguns poucos estudos atuais demonstram a aplicabilidade desse tipo de avaliação para a linguagem escrita, como o de Alavi e Taghizadeh (2014), que focou na avaliação dinâmica de escrita em aprendizes de uma segunda língua; ou o de Shrestha e Coffin (2012), na avaliação de escrita de estudantes de ensino superior, realizando a mediação por meio do computador e aprimorando o desempenho em escrita dos alunos. Há também o estudo de Xiaoxiao e Yan (2010) para avaliação e aprimoramento da aprendizagem em língua inglesa; e de verificação do impacto do treinamento, aliado à autoavaliação, em uma tarefa de escrita em língua inglesa, em universitários iranianos (Mazloomi & Kharibi, 2016). Poehner et al. (2015) também utilizaram um método computadorizado para avaliação dinâmica, sendo que o *feedback* dado nesses casos pode configurar-se como recurso contínuo e importante para a aprendizagem individualizada.

Os estudos citados demonstram a efetividade da avaliação dinâmica no aprimoramento do desempenho em escrita em diferentes contextos, avaliando o potencial de aprendizagem de cada sujeito, mediando o processo de avaliação e de aprendizagem. Todos os estudos chamam a atenção para a principal característica da avaliação dinâmica, que é avaliar promovendo a aprendizagem dos sujeitos. Ainda, demonstram que o ensino com mediação e diálogo amplia o interesse e a competência dos alunos no desenvolvimento de sua habilidade escrita (Shrestha & Coffin, 2012; Shrestha, 2013; Gurgel et al., 2017).

Por outro lado, ainda se observa dificuldade de escrita em universitários, sendo este um reflexo da aprendizagem precária realizada em contextos escolares prévios (Gebril & Plakans, 2014). É preciso, portanto, que sejam ampliadas as estratégias de avaliação no âmbito escolar, de crianças e adolescentes, a fim de que se possa avaliar de forma mais efetiva a habilidade de escrita desses sujeitos. Assim, estratégias de ensino personalizadas e efetivas ainda em momento escolar podem ser favorecidas,

preparando o sujeito para as demandas futuras (Gebril & Plakans, 2014).

Em específico, para escrever é preciso, inicialmente, articular uma ideia, fazendo uso do conteúdo lexical pré-existente, consciência fonológica e sintática, desenvolvimento pragmático, coesivo, realizando codificação e decodificação, além de decisões acerca da pontuação, regras ortográficas e conteúdo (Bindman et al., 2014; Salles & Correa, 2014). A narrativa, nesse processo de escrita, é uma etapa mais elaborada, sendo dependente também de fatores cognitivos e ambientais, variando em relação ao tema, contexto e intenções (Graham et al., 2013). Ainda, Nobile e Barrera (2016) e Gombert (1992) acrescentam que a consciência metatextual é uma habilidade inicial metalinguística imprescindível para o aprimoramento da habilidade de escrita narrativa, exigindo maior abstração.

A avaliação da habilidade de escrita narrativa é um ponto-chave para o acompanhamento do desenvolvimento e aprimoramento da escrita, especialmente em escolares. Daí a importância de se utilizar instrumentos que possam avaliar de forma fidedigna e válida o construto, apontando não somente os pontos fracos dos sujeitos, mas também seus pontos fortes e seu potencial de aprendizagem (Bartholomeu et al., 2015).

Estudos de validação de instrumentos de avaliação da escrita são escassos na literatura, especialmente se considerados aqueles construídos com base na metodologia dinâmica ou assistida. Por isso, Gebril e Plakans (2014) salientaram que é urgente o desenvolvimento de estratégias de avaliação que possam avaliar de forma precisa a produção textual de sujeitos. Além disso, a busca de evidências de validade dos instrumentos é essencial para designar a confiabilidade das pontuações e escores das escalas.

O TIDE (Joly & Schiavoni, 2013), em específico, tem como objetivo avaliar o potencial de aprendizagem em escrita narrativa de adolescentes. Já apresenta estudos de busca de evidências de validade (Joly et al., 2015; Gurgel et al., 2017). No estudo de Joly et al. (2015), objetivou-se buscar evidências de validade de estrutura interna e de variável externa para o TIDE em 127 escolares, do 6º e 9º ano, do interior de São Paulo. Joly e seus colaboradores, no citado estudo, demonstraram diferença estatisticamente significativa entre o número de palavras do módulo inicial para o final, além de correlações positivas e significativas entre o número de palavras com a nota em língua portuguesa para os dois módulos, confirmando-se evidências de validade do instrumento.

No estudo de Gurgel et al. (2017), objetivou-se apresentar o processo de busca de evidências de validade da estrutura interna do TIDE. Como resultado, observou-se concordância entre dois avaliadores que corrigiram o primeiro e o segundo textos dos participantes. Uma análise fatorial foi aplicada ao primeiro módulo (pré-teste), que revelou uma divisão em dois fatores e, ao segundo módulo (módulo instrucional), que foi

dividido em três fatores. A confiabilidade dos itens da TIDE foi verificada pelo alfa de Cronbach com coeficientes $> 0,7$. A análise do terceiro módulo (pós-teste) foi baseada no Teste de McNemar e mostrou resultados estatisticamente significativos que demonstraram uma evolução no desempenho dos participantes. Segundo Gurgel et al. (2017), é um instrumento válido e uma relevante ferramenta para a avaliação fonoaudiológica, psicológica e educacional, com natureza computacional para análises mais precisas do desempenho dos participantes e com menor custo.

Para além da teoria clássica de validação dos testes, a Teoria de Resposta ao Item tem se destacado como base para a busca de evidências de validade de instrumentos. Esta vem sendo utilizada, nos últimos anos, amplamente na Psicologia e na Fonoaudiologia, em estudos de análise psicométrica de instrumentos. Esse tipo de análise fornece informação sofisticada sobre os traços latentes estudados e a interpretação completa dos dados (Woodcock, 1999). A análise com base na TRI tem duas características principais: a primeira refere-se à possibilidade de observar a dificuldade dos itens e a habilidade dos sujeitos; a segunda é relativa à premissa de fornecer medidas invariantes do desempenho cognitivo, não dependendo dos itens que compõem a prova, nem das pessoas investigadas. Assim, a TRI permite a comparação dos resultados de provas de dificuldade variável e, ainda, o estudo da correlação entre os resultados na mesma escala de desempenho cognitivo. Essa metodologia vem sendo utilizada nos processos qualitativos da avaliação psicológica e educacional, em instrumentos de medida, como as escalas de habilidades, por exemplo, para avaliar e acompanhar o conhecimento adquirido pelos estudantes (Pasquali & Primi, 2003; Urbina, 2007; Valle, 2000; Vendramini, 2001).

Por meio da TRI também pode-se estimar a fidedignidade e os erros de mensuração, que são abordados do ponto de vista da função de informação de itens individuais do teste. É importante ressaltar que, quando se aplica a TRI para analisar desempenhos em instrumentos de avaliação de comportamentos, competências e atitudes, entende-se que esses elementos são o traço latente e que

estão relacionados à intensidade de adesão de cada item a esse construto, ou com que intensidade cada item avalia esse construto (Urbina, 2007). Na área da Psicologia, a TRI acompanha as tendências atuais voltadas para a valorização do uso de testes que sejam mais objetivos e breves. A TRI possibilita a otimização de instrumentos que, por exemplo, já foram validados por métodos clássicos, ampliando a qualidade dos diagnósticos (Sartes & Souza-Formigoni, 2013). A busca de evidências de validade com base na estrutura interna e a consideração da confiabilidade de um instrumento fornecem maior segurança e robustez para a aplicação dele, além da possibilidade de reprodutibilidade (Demarchi et al., 2019). Sendo assim, o objetivo do presente estudo é buscar evidências de precisão e de validade com base na estrutura interna do Teste Informatizado e Dinâmico de Escrita, utilizando-se da Teoria de Resposta ao Item.

Método

Participantes

O número de participantes foi calculado com base no critério item/sujeito (Hair, 2006). A amostra do presente trabalho foi composta por 304 participantes, escolhidos aleatoriamente em escolas públicas e privadas participantes da pesquisa. As escolas foram convidadas a participar por conveniência e estavam localizadas na região Central, Leste e Norte de Porto Alegre, RS, Brasil. Os participantes do estudo cursavam da quinta a oitava séries, sendo a maior parte da sétima série (37,1%). A idade dos participantes variou entre 10 e 17 anos, estando a maioria entre 12 e 13 anos de idade (43,2%). Quanto ao sexo, 64,5% da amostra era do sexo feminino. Quanto ao tipo da escola, apenas 8,2% eram provenientes da rede privada de ensino. A rede brasileira de educação consiste em educação infantil, educação primária e ensino médio. Quanto aos dados escolares, a maior parte da amostra iniciou a escola aos seis anos (48,8%), frequentou a educação infantil (82,1%) e 9,1% teve ou tem problemas para ler ou escrever de acordo com os pais e/ou cuidadores (que preencheram os questionários com informações sobre os participantes). Os dados sobre a amostra estão descritos na Tabela 1.

Tabela 1
Características da Amostra (N=304 Participantes)

	Características da Amostra	% (n)
Sexo		
	Feminino	64,5% (196)
	Masculino	35,5% (108)
Idade		
	10-11 anos	14,6% (44)
	12-13 anos	43,2% (131)
	14-15 anos	37,5% (114)
	16-17 anos	4,8% (15)

Tabela 1 (continuação)
 Características da Amostra (N=304 Participantes)

Características da Amostra	% (n)
Escolaridade	
Quinto	2% (7)
Sexto	33,4% (101)
Sétimo	37,1% (113)
Oitavo	27,4% (83)
Tipo da escola	
Pública	91,8% (279)
Privada	8,2% (25)
Performance escolar segundo os respondentes do questionário	
Regular	17,9% (55)
Bom	34,6% (105)
Muito bom	27,1% (82)
Excelente	20,4% (62)
Dificuldades escolares segundo pais/responsáveis	
Leitura	11,7% (35)
Escrita	14,6% (44)
Matemática	32,9% (100)
Atenção	33,9% (103)
Questões emocionais	10,4% (32)

Os critérios de seleção dos participantes foram: estar regularmente matriculado na quinta a oitava série e fornecer consentimento informado assinado pelos responsáveis legais. Os critérios de exclusão incluíram: problemas neurológicos relatados, defasagem escolar superior a dois anos, ausência de cooperação, não conclusão do teste por qualquer motivo ou outras condições físicas que impedissem a avaliação. Tais critérios de exclusão foram assim determinados por serem considerados como geradores de possibilidade de viés nas respostas dos sujeitos, especialmente considerando as questões escolares e de desenvolvimento da escrita necessárias para responder ao instrumento.

Instrumento

O objetivo do TIDE (Joly & Schiavoni, 2013) é avaliar o potencial de aprendizagem em escrita narrativa de adolescentes. Foi construído com base nos princípios da avaliação dinâmica do potencial de aprendizagem, de acordo com as premissas teóricas de Sternberg e Grigorenko (2002) e a estrutura teórica da produção escrita de Hayes e Flower (1980). O teste é constituído de três partes: um pré-teste, um módulo instrucional e um pós-teste. O sujeito passa por cada uma dessas etapas conforme responde às perguntas ou realiza as tarefas e, após isso, "clica" em seguir. No pré-teste, inicialmente, o participante deve preencher informações como nome, escolaridade e idade. Em seguida, o teste apresenta instruções sobre a tarefa proposta: escrever um texto narrativo sobre o tema "Herói por um dia". Posteriormente,

o participante é solicitado a ler o texto e corrigir o que achar necessário. Não há sugestões sobre o desempenho do respondente nessa etapa. O módulo pré-teste é composto por sete itens.

O módulo instrucional é composto por 19 itens (nomeados de 1.1 a 7.1), com perguntas oferecidas aos participantes, que são disponibilizadas logo após o sujeito ter feito a leitura citada anteriormente de seu texto. Supõe que as estratégias oferecidas facilitam a aprendizagem da escrita de um texto narrativo e se baseia nos processos de Hayes e Flower (1980). Esse modelo de composição escrita não descreve fases sucessivas, mas os processos envolvidos. Um desses processos é o planejamento, que serve para extrair informações sobre a natureza da tarefa e sobre a memória de longo prazo, e usá-lo para estabelecer um plano que orienta a produção de um texto, incluindo três subgrupos: produção de ideias, organização e definição de metas. O segundo processo é a tradução, que se destina a transformar o planejamento da mensagem em frases escritas, que devem cumprir as normas da linguagem escrita. A avaliação do texto produzido e a melhoria da sua qualidade correspondem à revisão, o terceiro processo de composição da escrita. Este último contém dois subprocessos: leitura e edição.

O indivíduo recebe instruções sobre como escrever uma narrativa e os elementos necessários para formulá-la, relacionados às personagens, ao cenário, à situação e ao problema. Algumas perguntas são feitas a fim de que o respondente reflita sobre a sua própria construção escrita. Espera-se que ele responda às perguntas e complete o

texto, se houver algum elemento faltando, de acordo com o seu próprio julgamento. As primeiras perguntas se concentram nas personagens da história, como "Quem são as personagens principais?", "Como elas são?". Em seguida, são feitas perguntas sobre o cenário, situação ou problema enfrentado, resposta ao problema, ação, solução e reação das personagens da história. Finalmente, pede-se ao participante que revise seu texto com base nas perguntas finais. Então, ele deve rever se as frases estão completas, se os nomes próprios começam com letras maiúsculas, se as palavras estão escritas de forma adequada e se os sinais de pontuação foram usados corretamente. O resultado é analisado em uma escala Likert de três pontos (não adicionou a resposta à pergunta no texto, adicionou parcialmente a resposta à pergunta no texto, adicionou a resposta à pergunta no texto). Cada elemento da narrativa tem uma cor diferente, de modo que fica claro para o aluno e para o avaliador o quanto ele adiciona ou altera em cada um desses elementos da narrativa.

No módulo pós-teste, que surge logo após o sujeito ter finalizado as respostas no módulo instrucional, o participante analisa sua produção textual como um todo. Ele tem a oportunidade de reler e editar o texto, se necessário, sem a interferência ou dica do avaliador. Como é um teste computadorizado, o participante pode alterar e acessar qualquer item no módulo que achar necessário, quantas vezes for necessário. As alterações feitas ao editar o texto aparecem na tela em uma cor diferente.

O avaliador analisa comparativamente o texto inicial e o texto reformulado, de forma a verificar o desempenho do participante após as instruções fornecidas. Nessa fase, dois avaliadores, um fonoaudiólogo e um psicólogo, pós-graduados e rigorosamente treinados, avaliam os textos. Quatro aspectos específicos são observados no texto de cada participante, considerando a intervenção, ocorrendo ao longo do módulo instrucional, sendo padronizada e a mesma para todas as escolas envolvidas e grupos de alunos participantes (inicial – pré-intervenção e final – após intervenção): mudanças em todo o texto em relação ao contexto do tema, em relação ao desenvolvimento do texto (situação ou problema/solução/ação proposta para resolver/a reação do problema), tema e conteúdo geral. Os textos foram avaliados com base em uma escala Likert de três pontos (não cumprem o critério, cumprem parcialmente o critério, cumprem plenamente o critério). A diferença entre a pontuação do texto final e a pontuação do texto inicial foi avaliada para cada participante. Nessa última fase, é possível medir a zona de desenvolvimento proximal do participante, ou seja, a diferença entre o desempenho potencial e real de cada participante (Vygotsky, 1988).

Procedimentos

O presente trabalho foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Ciências da Saúde de Porto Alegre, sob o protocolo número 311.451.

Os pais ou cuidadores responsáveis pelos alunos das escolas participantes foram contatados para a assinatura do Termo de Consentimento Livre e Esclarecido. Participaram da pesquisa apenas os alunos cujos pais ou cuidadores responsáveis deram o seu consentimento informado. Para a aplicação do instrumento, foram organizados grupos de até três alunos, monitorados por três pesquisadores, de forma simultânea. Os grupos foram organizados por idade e nível de escolaridade. Os dados foram coletados em uma sala privada de cada escola, silente, sem interferência externa, durante o horário regular de aula. Os alunos receberam as instruções por um dos pesquisadores do presente estudo e fizeram o teste individualmente, cada um em um computador, numa mesma sala, utilizando fones de ouvido para evitar interferência no teste dos outros respondentes. Foram utilizados três *notebooks*, pertencentes aos pesquisadores, com capacidade básica de *hardware* para suportar o *software* TIDE. A aplicação completa do TIDE (incluindo pré-teste, modelo de instrução e pós-teste) levou 60 minutos em média.

O *software* TIDE armazena automaticamente os testes em um banco de dados, em uma pasta específica no computador. Ressalta-se que esta pesquisa não envolveu riscos para os participantes e que a privacidade dos sujeitos foi preservada, conforme normativas éticas relacionadas a pesquisas envolvendo seres humanos. A aprovação ética foi obtida pelo Comitê de Ética em Pesquisa da Fundação Universidade Federal de Ciências da Saúde de Porto Alegre, Brasil, sob protocolo 502.515

Análise dos Dados

As respostas aos testes foram corrigidas e pontuadas de acordo com os critérios pré-definidos para o TIDE, descritos anteriormente, por dois avaliadores independentes, rigorosamente treinados para a aplicação e interpretação dos dados. O referido treinamento foi realizado ao longo de um semestre, acontecendo em paralelo ao delineamento da coleta de dados da presente pesquisa. O treinamento envolveu aprofundamento teórico-prático nas premissas do TIDE, nas teorias de buscas de evidências de validade, além do desenvolvimento e avaliação da linguagem escrita, especialmente em sua dimensão narrativa. As divergências na correção foram resolvidas por consenso e análise conjunta.

A fim de buscar evidências de validade com base na estrutura interna do TIDE, realizou-se uma análise dos módulos pré-teste e instrumental do teste. Buscou-se inicialmente evidências de unidimensionalidade, pré-requisito para análises da TRI e após estimar os parâmetros dos itens e o nível de habilidade dos participantes. Utilizou-se o *software* R (R Core Team, 2018) para análise dos dados. Para o módulo pré-teste, optou-se pelo modelo de dois parâmetros, por apresentar melhor ajuste. O ajuste dos modelos foi avaliado pelos resíduos padronizados. Para o módulo instrucional, utilizou-se o modelo politômico de resposta gradual (Samejima, 1969).

Considerando as premissas da TRI, para o módulo pré-teste, foi adotado valor de discriminação de pelo menos 1 para consideração de um bom ajuste dos itens (Linacre, 2015). Considerando o módulo instrucional, para o modelo de resposta gradual de Samejima, necessita-se valores de discriminação entre 0,70 e 4,0. Em relação à dificuldade do item para responder de forma afirmativa, parâmetro b , este deve variar entre -4 e $+4$ (Pasquali & Primi, 2003; Sartes & Souza-Formigoni, 2013). Por fim, para o coeficiente alfa de Cronbach, na avaliação da fidedignidade e precisão (CFP, 2003; Cronbach, 1996).

Resultados

Inicialmente, avaliou-se a dimensionalidade dos módulos em análise por meio de análise fatorial. Em relação ao pré-teste, verificou-se que a primeira raiz característica (autovalor) é superior às demais, estando associada a 41,62% da variabilidade, indicando que o primeiro fator é responsável por mais de 40% da variabilidade. Dessa forma, o módulo pré-teste do instrumento é unidimensional, já que tem um fator preponderante, que é assim considerado sempre que o primeiro fator explicar ao menos 20% (Reckase, 1979). Para os itens do módulo instrucional, verificou-se que a primeira raiz característica (autovalor) é muito maior que as demais, estando associada a 36,55% da variabilidade. Sendo assim, o conjunto de itens do módulo instrucional também satisfizeram condição de unidimensionalidade para análises estatísticas com base na TRI.

Satisfeitas as condições para a realização da TRI, para o módulo pré-teste, foi ajustado o modelo Logístico de dois parâmetros (ML2) para calibração dos itens. Foi observado um coeficiente alfa de Cronbach de 0,82 para

o conjunto de sete itens que mensuram o construto escrita narrativa, indicando alta fidedignidade e precisão desse instrumento.

Para o módulo pré-teste, foi adotado valor de discriminação de pelo menos 1 para consideração de um bom ajuste dos itens (Linacre, 2015). Nesse módulo pré-teste, todos os itens apresentaram valores adequados de discriminação, que variaram entre 1,88 (item “personagens”) e 3,49 (item “cenário”). A dificuldade dos itens variou de $-0,01$ a $1,27$, o que sugere que o teste avalia apropriadamente estudantes em diferentes níveis da habilidade estimada. Observa-se, no entanto, a falta de itens muito fáceis (com dificuldade abaixo de -2 , por exemplo), o que demonstra a complexidade da tarefa solicitada (Tabela 2). Sendo assim, há maior discriminação para níveis médios de habilidade.

Na avaliação dos itens do módulo instrucional, o coeficiente alfa de Cronbach foi de 0,91 indicando, assim como no módulo pré-teste, alta fidedignidade e precisão, além de adequação para a avaliação do construto que pretende avaliar. Para a análise do módulo instrucional, utilizou-se o modelo politômico de resposta gradual (Samejima, 1969). O modelo ajustado forneceu valores de discriminação considerados adequados para todos os itens ($>1,6$, z value=21,164) e diferentes níveis de dificuldade dos itens. Esses valores não representam pouca ou excessiva discriminação, revelando que os itens mensuram adequadamente a probabilidade de resposta afirmativa a cada um deles. Já o parâmetro dificuldade dos itens apresenta valores diferentes, que variaram entre $-1,41$ a $0,22$ para *extrmt1* e entre $-0,55$ a $0,91$ para *extrmt2*, sugerindo que o teste avalia apropriadamente estudantes com diferentes níveis da habilidade (Tabela 3)

Tabela 2
Dados de Dificuldade e Discriminação de Cada Item do Módulo Pré-Teste

	Dificuldade	Discriminação
Personagens	-0,016	1,884
Cenário	0,282	3,495
Situação	0,673	2,745
Resposta	0,952	2,910
Ação	1,149	2,703
Solução	1,272	2,403
Reação	0,555	2,755

Tabela 3
Informação Estatística Relativa ao Modelo de Resposta Gradual de Samejima Referente ao Módulo Instrucional do TIDE

Item	Extrmt1	Extrmt2	Discriminação
1.1	-1,411	-0,552	1,843
1.2	0,135	0,916	1,843
1.3	0,222	0,608	1,843
1.4	-0,384	0,494	1,843

Tabela 3 (continuação)

Informação Estatística Relativa ao Modelo de Resposta Gradual de Samejima Referente ao Módulo Instrucional do TIDE

Item	Extrmt1	Extrmt2	Discriminação
1.5	-0,648	0,148	1,843
1.6	-0,618	-0,019	1,843
1.7	-0,100	0,372	1,843
1.8	-0,277	0,434	1,843
2.1	0,099	0,770	1,843
2.2	-0,822	-0,064	1,843
2.3	0,355	1,035	1,843
3.1	-0,897	0,049	1,843
4.1	-0,796	0,187	1,843
5.1	-0,674	-0,041	1,843
5.2	-0,049	0,507	1,843
6.1	-0,664	0,026	1,843
6.2	-0,997	-0,294	1,843
6.3	-0,310	0,355	1,843
7.1	-0,200	0,638	1,843

Discussão

O Teste Informatizado e Dinâmico de Escrita objetiva avaliar o potencial de aprendizagem em escrita narrativa de adolescentes, configurando-se como um instrumento inovador na área (Gurgel et al., 2017). Conforme observado nos resultados do presente estudo, fez-se necessária esta extensa análise de suas características psicométricas, a fim de verificar as suas evidências de validade.

Inicialmente, a análise da unidimensionalidade indicou possibilidade de uso da Teoria de Resposta ao Item para análise do instrumento em questão. Ainda, os resultados apresentados demonstram que os itens avaliam, de fato, o construto proposto, o potencial de aprendizagem em escrita de textos narrativos. Além de alfas elevados, os parâmetros de discriminação e dificuldade de cada item, analisados por meio da TRI, sugerem que os itens são fidedignos e válidos para mensurar o construto.

Em relação ao módulo pré-teste, sabe-se que a discriminação dos itens corresponde ao parâmetro a , e deve ser entendida como a capacidade do item de diferenciar indivíduos quanto à intensidade do traço latente, devendo apresentar-se com valor positivo (Pasquali & Primi, 2003; Sartes & Souza-Formigoni, 2013). O parâmetro de discriminação corresponde ao ângulo de inclinação da curva, sendo as mais inclinadas compatíveis com maior discriminação. Portanto, quanto maior o valor do parâmetro de discriminação, mais o item detecta diferenças entre os respondentes (Sartes & Souza-Formigoni, 2013). Nesse sentido, observa-se que todos os itens se encontram dentro dos valores esperados para os parâmetros de discriminação e de dificuldade, sendo o item relacionado ao cenário aquele com maior discriminação. O item “personagem” é o que menos discrimina os participantes, e os demais itens apresentam valores semelhantes para

esse parâmetro. Apesar de os demais itens apresentarem valores semelhantes de discriminação, são diferentes em relação ao conteúdo e abordam temáticas diversas relacionadas ao texto narrativo produzido. Sendo assim, não se sugere a retirada de nenhum destes. Considerando a escrita de textos narrativos, pressupõe-se que o item “cenário” tenha uma influência mais ampla sobre a construção narrativa, sendo portanto mais discriminativo. Ao passo que o item “personagens” pode ser um dos primeiros pontos criados da história, sendo menos suscetível a alterações ao longo do processo instrucional do teste, sendo portanto menos discriminativo.

Em relação à dificuldade do item para responder de forma afirmativa, parâmetro b , este deve variar entre -4 e $+4$ (Pasquali & Primi, 2003; Sartes & Souza-Formigoni, 2013). Os itens mais difíceis no módulo pré-teste são “resposta”, “situação” e “reação”, todos relacionados com pontos centrais de desenvolvimento da narrativa e de encadeamento da história produzida. O item mais fácil é o item “personagem”, sendo este dado compatível com a teoria, uma vez que esse item é mais descritivo, além de essencial para a construção narrativa.

Em relação ao módulo instrucional, sabe-se que o modelo de resposta gradual de Samejima utilizado permite obter mais informação por meio das respostas dos indivíduos, necessitando apresentar valores de discriminação entre 0,70 e 4,0. Esse tipo de análise fornece um valor único para o parâmetro a (discriminação), e diferentes valores para o parâmetro correspondente à dificuldade do item (podendo ser apresentada a média desses valores (Sartes & Souza-Formigoni, 2013)). Para a amostra deste estudo, também se observa adequação dos valores apresentados.

O presente estudo atendeu aos objetivos propostos, demonstrando evidências de validade com base na

estrutura interna de um novo instrumento para avaliação do potencial de aprendizagem em escrita narrativa de adolescentes. Observou-se que os itens apresentam adequados valores de discriminação e dificuldade, analisados com base na Teoria de Resposta ao Item, formando um conjunto de itens com bom índice de precisão.

Como limitações do instrumento apresentado, pode-se considerar a falta de flexibilidade nas instruções fornecidas ao longo da realização do teste, uma vez que a ferramenta computadorizada é pré-programada. Assim, as instruções ao longo do teste não apresentam variação (Poehner et al., 2015; Alderson, 2007). Ainda como limitação, ressalta-se que o parâmetro c , da Teoria de Resposta ao Item, relacionado à probabilidade de acerto ao acaso de cada item, não foi avaliado, em virtude da escolha dos modelos de TRI mais adequados à estrutura dos módulos do teste.

Finalmente, este estudo é um pioneiro na área e, além de apresentar uma ferramenta inovadora de evidência sobre a validade estrutural interna, também estimula pesquisas futuras sobre o tema. Atualmente, estão sendo realizados outros estudos considerando a busca de evidências de validade do TIDE, pelo presente grupo de pesquisa, de forma a possibilitar futuramente análises comparativas sobre o desempenho do instrumento em populações distintas e associado a outros instrumentos.

Seguindo as sugestões de Muniz, Seabra e Primi (2015), são necessários estudos para investigar a relação entre a avaliação do potencial de aprendizagem e construções relacionadas, como o desempenho educacional dos participantes e a situação escolar. O instrumento apresentado neste estudo corrobora os pressupostos teóricos de Vygotsky e outros autores da área, como Poehner et al. (2015), que enfatizam o alinhamento entre a avaliação e a mediação, favorecendo, assim, o diagnóstico mais detalhado das potencialidades e dificuldades dos avaliados, direcionando o planejamento de ensino e/ou de tratamento.

Agradecimentos

Agradecemos a Prof. Dra. Maria Cristina Rodrigues Azevedo Joly por todo seu auxílio para a realização dessa pesquisa. Agradecemos também à CAPES e ao CNPq pelo fomento à pesquisa.

Financiamento

Todas as fontes de financiamento para elaboração e produção do estudo (coleta, análise e interpretação dos dados, bem como, escrita dos resultados no presente no manuscrito) foram fornecidas pelo projeto de pesquisa 'Bolsa de Doutorado CAPES – Demanda Social.

Contribuições dos autores

Todos os autores participaram da elaboração do manuscrito. Especificamente, as autoras Léia Gurgel, Livia Teixeira e Caroline Reppold participaram da redação inicial do estudo – conceitualização, investigação, visualização; as autoras Léia Gurgel, Livia Teixeira, Caroline Reppold, Vanessa Kaiser e Monica de Oliveira participaram da análise dos dados, e as autoras Léia Gurgel, Livia Teixeira, Caroline Reppold, Vanessa Kaiser e Monica de Oliveira participaram) da redação final do trabalho, revisão e edição.

Disponibilidade dos dados e materiais

Todos os dados e sintaxes gerados e analisados durante esta pesquisa serão tratados com total sigilo devido às exigências do Comitê de Ética em Pesquisa em Seres Humanos. Porém, o conjunto de dados e sintaxes que apoiam as conclusões deste artigo estão disponíveis mediante razoável solicitação ao autor principal do estudo.

Conflito de interesses

Os autores declaram que não há conflitos de interesses.

Referências

- Alavi, S. M., & Taghizadeh, M. (2014). Dynamic assessment of writing: The impact of implicit/explicit mediations on L2 learners' internalization of writing skills and strategies. *Educational Assessment*, 19(1), 1-16. <https://doi.org/10.1080/10627197.2014.869446>
- Alderson, J. C. (2007). The challenge of (diagnostic) testing: Do we know what we are measuring? Em J. Fox, M. Wesche, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 21-39). Ottawa: University of Ottawa
- Bartholomeu, D., Montiel, J. M., Neia, S., Fiamenghi, G. A., & Silva, M. C. (2015). School Performance in Writing, Reading and Mathematics Related to Social Skills. *International Journal of Psychology and Behavioral Sciences*, 5(1): 1-5. <https://doi.org/10.5923/j.ijpbs.20150501.01>
- Bindman, S. W., Skibbe, L. E., Hindman, A. H., Aram, D., & Morrison, F. J. (2014). Parental writing support and preschoolers' early literacy, language, and fine motor skills. *Early childhood research quarterly*, 29(4), 614-624. <https://doi.org/10.1016/j.ecresq.2014.07.002>
- Conselho Federal de Psicologia – CFP (2003). *Resolução nº 002/2003* [On-line]. <https://site.cfp.org.br/wp-content/uploads/2003/03/formul%C3%A1rio-anexo-res-02-03.pdf>
- Cronbach, L. J. (1996). *Fundamentos da Testagem Psicológica*. Porto Alegre: Artmed.
- Demarchi, G. S. D. S., Andrade, M. D., Novelli, M. M. P. C., Katz, N., & Uchôa-Figueiredo, L. D. R. (2019). Análise da consistência interna da versão em português da Avaliação Cognitiva Dinâmica de Terapia Ocupacional para Crianças (DOTCA-Ch) em estudantes de 06 a 12 anos. *Cadernos Brasileiros de Terapia Ocupacional*, (AHEAD).

- Enumo, S. R. F. (2005). Avaliação assistida para crianças com necessidades educacionais especiais: Um recurso auxiliar na inclusão escolar. *Revista Brasileira de Educação Especial*, 11(3): 335-354. <https://doi.org/10.1590/S1413-65382005000300003>
- Ferrioli, S. H. T., Linhares, M. B. M., Loureiro, S. R., & Marturano, E. M. (2001). Indicadores de potencial de aprendizagem obtidos através da avaliação assistida. *Psicologia: Reflexão e Crítica*, 14(1), 35-43. <https://doi.org/10.1590/S0102-79722001000100003>
- Gebril, A., & Plakans, L. (2014). Assembling validity evidence for assessing academic writing: Rater reactions to integrated tasks. *Assessing Writing*, 21(2014), 56-73. <https://doi.org/10.1016/j.asw.2014.03.002>
- Gombert, J. E. (1992). *Metalinguistic development*. University of Chicago Press.
- Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing*, 26(1), 1-15. <https://doi.org/10.1007/s11145-012-9395-2>
- Gurgel, L. G., de Oliveira, M., Joly, M. C., & Reppold, C. T. (2017). Learning Potential in Narrative Writing: Measuring the Psychometric Properties of an Assessment Tool. *Frontiers in psychology*, 10(8), 719. <https://doi.org/10.3389/fpsyg.2017.00719>
- Hair Jr, J. F. (2006). Black, WC/Babin, BJ/Anderson, RE & Tatham, RL (2006): Multivariate Data Analysis. *Auflage, Upper Saddle River*.
- Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of writing processes. Em L. W. Gregg & E. R. Steinberg (Eds.). *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Joly, M. C. R. A., & Schiavoni, A. (2013). *Teste Informatizado e Dinâmico de Escrita*. [Projeto de pesquisa não publicado], Programa de Pós-graduação em Processos de Desenvolvimento e Aprendizagem, Universidade de Brasília, Distrito Federal, Brasil.
- Joly, M. C. R. A., Schiavoni, A., Agostinho, A., & Dias, A. S. (2015). Avaliação dinâmica e produção textual: Evidências de validade para o TIDE no fundamental II. *Cadernos de Pós-Graduação em Distúrbios do Desenvolvimento*, 15(1), 50-62. <https://biblat.unam.mx/hevila/Cadernosdeposgraduacaoemdisturbiosdodesenvolvimento/2015/vol15/no1/5.pdf>
- Linacre J. M. (2015). *A user's guide to Winsteps Ministep: Raschmodel computer programs*. <http://www.winsteps.com>
- Mazloomi, S., & Khabiri, M. (2016). Diagnostic Assessment of Writing through Dynamic Self-Assessment. *International Journal of English Linguistics*, 6(6), 19-31. <https://doi.org/10.5539/ijel.v6n6p19>
- Muniz, M., Seabra, A. G., & Primi, R. (2015). A relação entre potencial de aprendizagem e desempenho acadêmico: Predição pelo Teste Dinâmico Informatizado de Raciocínio Indutivo para Crianças. *Estudos de Psicologia*, 32(3), 343-356. <https://doi.org/10.1590/0103-166X2015000300001>
- Nobile, G. G., & Barrera, S. D. (2016). Desempenho ortográfico e habilidades de produção textual em diferentes condições de solicitação. *Psicologia: Teoria e Pesquisa*, 32(2), 1 e32226. <http://dx.doi.org/10.1590/0102-3772e32226>
- Pasquali, L., & Primi, R. (2003). Fundamentos da Teoria de Resposta ao Item – TRI. *Avaliação Psicológica*, 2(2), 99-110. <http://pepsic.bvsalud.org/pdf/avp/v2n2/v2n2a02.pdf>
- Poehner, M. E., Zhang, J., & Lu, X. (2015). Computerized dynamic assessment (C-DA): Diagnosing L2 development according to learner responsiveness to mediation. *Language Testing*, 32(3), 337-357. <https://doi.org/10.1177/0265532214560390>
- R Development Core Team. (2018). *R: A language an environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Software. <https://www.R-project.org>
- Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of educational statistics*, 4(3), 207-230. <https://doi.org/10.2307/1164671>
- Salles, J. F. D., & Correa, J. (2014). A produção escrita de histórias por crianças e sua relação com as habilidades de leitura e escrita de palavras/pseudopalavras. *Psicologia USP*, 25(2), 189-200. <https://doi.org/10.1590/0103-6564A20133813>
- Samejima, F. A. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph*, 17. <https://doi.org/10.1002/j.2333-8504.1968.tb00153.x>
- Sartes, L. M. A., & Souza-Formigoni, M. L. O. D. (2013). Avanços na psicometria: Da Teoria Clássica dos Testes à Teoria de Resposta ao Item. *Psicologia: Reflexão e Crítica*, 26(2), 241-250. <http://dx.doi.org/10.1590/S0102-79722013000200004>
- Shrestha, P. (2013). Dynamic assessment and academic writing: Evidence of learning transfer? Em M. Kavanagh & L. Robinson (Eds.). *Janus moment in EAP: Revisiting the past and building the future – Proceedings of the 2013 BALEAP Conference* (pp.19-21), University of Nottingham.
- Shrestha, P., & Coffin, C. (2012). Dynamic assessment, tutor mediation and academic writing development. *Assessing Writing*, 17(1), 55-70. <http://dx.doi.org/10.1016/j.asw.2011.11.003>
- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. Cambridge university press.
- Tzuriel, D. (2000). Dynamic assessment of young children: Educational and intervention perspectives. *Educational Psychology Review*, 12(4): 385-435. <https://doi.org/10.1023/A:1009032414088>
- Urbina, S. (2007). *Fundamentos da testagem psicológica*. Porto Alegre: Artmed.
- Valle, R. C. (2000). Teoria de Resposta ao Item. *Estudos em Avaliação educacional*, 21(2000),7-91. <https://doi.org/10.18222/eaec02120002225>
- Vendramini, C. M. M. (2001). Aplicação da Teoria de Resposta ao Item na Avaliação Educacional. Em R. Primi (Ed.), *Temas em Avaliação Psicológica* (pp. 116-130). Campinas: Instituto Brasileiro de Avaliação Psicológica.
- Vygotsky, L. S. (1988). Aprendizagem e desenvolvimento intelectual na idade escolar. Em L. S. Vigotsky, A. R. Luria, & A. N. Leontiev (Orgs), *Linguagem, desenvolvimento e aprendizagem* (pp. 103-117). São Paulo: Ícone: EDUSP.
- Woodcock, R. W. (1999). What can Rasch-Based scores convey about a person's test performance? Em S. Embretson & S. Hershberger (Eds.), *The new rules of measurement: What every psychology educator should know* (105-127). Mahwah, NJ: Lawrence Erlbaum.
- Wright, B. D. (1999). Fundamental measurement for psychology. Em S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 65-104). Mahwah, NJ: Lawrence Erlbaum.

Xiao-xiao, L., & Yan, L. (2010). A case study of dynamic assessment in EFL process writing. *Chinese Journal of Applied Linguistics*, 33(1), 24-40. <https://www.semanticscholar.org/paper/A-Case-Study-of-Dynamic-Assessment-in-EFL-Process-Xiao-xiao-Yan/bb652ddef3be4bc68804e71ebfbaadc7282cbcab#paper-header>

recebido em setembro de 2019
aprovado em junho de 2020

Sobre as autoras

Léia Gonçalves Gurgel é Fonoaudióloga (UFCSPA), doutora em Ciências da Saúde (UFCSPA). Atualmente é Fonoaudióloga Clínica do Departamento de Fonoaudiologia da UFSC, Campus Florianópolis.

Livia Padilha de Teixeira é Psicóloga, Mestre em Ciências da Saúde (UFCSPA). Atualmente é doutoranda do Programa de Pós-graduação em Psicologia da UFRGS.

Vanessa Kaiser é Psicóloga, mestre em Ciências da Saúde (UFCSPA). Atualmente é doutoranda do PPG em Ciências da Saúde da Universidade Federal de Ciências da Saúde de Porto Alegre, RS.

Monica Maria Celestina de Oliveira é Estatística (Universidade Federal da Bahia), doutora em em Epidemiologia (UFRGS). Atualmente é professora adjunta do Departamento de Saúde Coletiva da UFCSPA, RS.

Caroline Tozzi Reppold é Psicóloga (UFRGS), Doutora em Psicologia (UFRGS). Atualmente é Professora adjunta do Departamento de Psicologia da UFCSPA, RS.

Como citar este artigo

Gurgel, L. G., Teixeira, L. P., Kaiser, V., Oliveira, M. M. C., & Reppold, C. T. (2021). Teste informatizado e dinâmico de escrito: Evidências de validade baseadas na TRI. *Avaliação Psicológica*, 20(1), 33-42. <http://dx.doi.org/10.15689/ap.2021.2001.18908.04>