

# Estudos de validade entre instrumentos que avaliam habilidades linguísticas

## *Studies of validity between instruments to evaluate reading and writing abilities*

Neide de Brito **CUNHA**<sup>1</sup>

Acácia Aparecida Angeli dos **SANTOS**<sup>2</sup>

### Resumo

Medidas de construtos psicológicos tornam-se mais eficazes à medida que acumulam mais evidências de validade. Nessa perspectiva, este estudo teve por objetivo buscar evidências de validade convergente entre a Escala de Avaliação de Dificuldades na Aprendizagem da Escrita, a Escala de Reconhecimento de Palavras e o Teste de Cloze. Foram averiguadas as habilidades linguísticas e investigada a variável série para buscar derivar evidências de validade de critério. Participaram 266 crianças, entre 8 e 13 anos de idade, de ambos os sexos, de terceiras e quartas séries do ensino fundamental de escolas públicas, particulares e do Sesi, do interior de São Paulo. Os resultados permitiram encontrar evidência de validade de critério pela série somente no Reconhecimento de Palavras. A análise da correlação entre as médias do Reconhecimento de Palavras e o Cloze, pelas categorias da Escala de Avaliação de Dificuldades na Aprendizagem da Escrita, também indicou índices altamente significativos, permitindo encontrar evidência de validade convergente entre os instrumentos.

**Unitermos:** Compreensão da leitura. Escrita. Distúrbios de aprendizagem. Psicometria. Teste de Cloze.

### Abstract

*The aim of this study was to search for evidence of convergent validity between Evaluation Scale of Difficulties in Learning to Write, the Word Recognition Scale and the Cloze Test. Linguistic abilities were assessed using these specific instruments and other variables were also investigated, from the school grade to the recognition of evidence of criterion validity. In addition, the quality of children's mistakes with regard to the Cloze test was analyzed according to the Evaluation Scale of Difficulties in Learning to Write, which classified the children into groups with and without learning difficulties. The sample was composed of 266 elementary school children in public, private and Social Service for Industry schools in São Paulo. The results only permitted us to find evidence of criterion validity with Word Recognition at the grade level. The analysis of correlation between the means of Word Recognition and the Cloze test, using the Evaluation Scale of Difficulties in Learning to Write categories, also indicated significant differences, suggesting convergent validity amongst the instruments.*

**Uniterms:** Reading comprehension. Handwriting. Learning disabilities. Psychometrics. Cloze Testing.

Um teste é um procedimento sistemático para comparar o comportamento de duas ou mais pessoas. A invenção dos testes psicológicos ocorreu no meio do

século XIX, quando a ênfase dos psicólogos era dada ao estudo exaustivo dos fenômenos mentais em poucos participantes. Posteriormente, houve um avanço dos

▼▼▼▼▼

<sup>1</sup> Universidade São Francisco, Curso de Psicologia. Itatiba, SP, Brasil.

<sup>2</sup> Universidade São Francisco, Programa de Pós-Graduação em Psicologia. R. Alexandre Rodrigues Barbosa, 45, 13251-040, Itatiba, SP, Brasil. Correspondência para/Correspondence to: A.A.A. SANTOS. E-mail: <acacia.santos@saofrancisco.edu.br>.

conceitos estatísticos, que começaram a sustentar que os erros de mensuração poderiam ser cancelados com o efeito massivo de grandes amostras. A tecnologia também avançou na aplicação dos testes com o advento do computador. A psicometria passou a ser um ramo da ciência estatística e a administração, a contagem e a interpretação dos testes foram objetivamente incrementadas e automatizadas (Cronbach, 1960; Goldstein & Hersen, 2000).

De acordo com Ramsay e Reynolds (2000), o desenvolvimento de um teste científico demanda perseverança e rigor, visto que ele tem que emergir da literatura científica da área de interesse e estar apoiado em sólida fundamentação teórica. Além disso, ele tem que ser submetido a um processo exaustivo de desenvolvimento e validação para assegurar que pode alcançar o objetivo a que se destina. Os usuários dos testes devem compará-los com outras medidas existentes para determinar seu mérito relativo.

Os psicólogos e outros profissionais recolhem amostras de conduta porque a partir delas podem fazer inferências fundadas sobre a conduta e funcionamento cognitivo das pessoas avaliadas. A primeira condição para que um teste sirva de base para realizar inferências de interesse é que a amostra de conduta recolhida seja precisa, ou seja, que os erros cometidos na medição sejam aceitáveis, já que nenhuma medição está totalmente isenta de erro (Muñiz, 2004).

A testagem psicológica vive um momento de mudanças rápidas no mundo todo. Existem orientações novas, um constante fluxo de testes novos, formas revisadas de testes antigos e dados adicionais que podem aperfeiçoar ou alterar a interpretação dos escores de testes existentes. Uma cobertura mais extensa de instrumentos de testagem e problemas em áreas especiais pode ser encontrada em livros que tratam do uso dos testes em campos como aconselhamento, prática clínica, seleção de pessoal e educação. Para acompanhar os desenvolvimentos atuais, as pessoas que trabalham com testes precisam estar familiarizadas com fontes mais diretas de informações sobre eles, sendo uma das mais importantes delas o *Mental Measurements Yearbook* (Anastasi & Urbina, 2000; Spies & Plake, 2005).

A psicometria teve grandes avanços em todos os ramos, inclusive na determinação do conceito de validade, desde que foram publicados os padrões da

American Educational Research Association (AERA) (1985), *American Psychological Association* (APA) e *National Council on Measurement in Education* (NCME). Porém, as três vias clássicas para a coleta de dados, a validade de conteúdo, de critério e de construto ainda são férteis (Muñiz, 2004).

Sisto, Codenotti, Costa e Nascimento (1979) verificaram nos catálogos de testes disponíveis a quase inexistência de testes brasileiros. Embora seja recente a preocupação em utilizar instrumentos de avaliação cujas propriedades psicométricas tenham sido estabelecidas para populações específicas, os autores já constataram que essa carência era suprida com a utilização de instrumentos estrangeiros. Foi levantado por eles o problema de que muitos dos testes aqui utilizados tinham sido apenas adaptados, sem que houvesse uma análise da conservação ou não das propriedades psicométricas obtidas na construção do instrumento em seu país de origem.

Estudos realizados posteriormente continuam a apontar a precariedade dos instrumentos psicológicos utilizados no Brasil (Noronha, 2002; Noronha, Freitas & Ottati, 2001; Noronha, Freitas, Baldo, Barbini & Almeida, 2004; Noronha, Oliveira & Beraldo, 2003; Vendramini & Noronha, 2002). Essas constatações associadas à maior consciência social dos profissionais da psicologia mobilizaram alguns segmentos da comunidade científica e profissional, levando à criação do Instituto Brasileiro de Avaliação Psicológica (IBAP) e de um programa de pós-graduação *Stricto Sensu* com área de concentração em avaliação psicológica, recomendado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), em 2001.

O Conselho Federal de Psicologia (CFP), refletindo toda a mobilização em torno da questão, estabeleceu através da Resolução nº 002/2003, a regulamentação do uso, elaboração e comercialização de testes psicológicos no Brasil. Para o caso específico de instrumentos de avaliação psicológica que utilizam questões de múltipla escolha ou similares, como inventários e escalas, a resolução indica como requisitos mínimos e obrigatórios: 1) fundamentação teórica com ênfase na definição do construto; 2) apresentação de evidências empíricas de validade e precisão das interpretações propostas para os escores do teste; 3) apresentação de dados empíricos sobre as propriedades psicométricas

dos itens do instrumento; 4) apresentação do sistema de correção e interpretação dos escores; 5) apresentação clara dos procedimentos de aplicação e correção, bem como as condições nas quais o teste deve ser aplicado; e 6) compilação das informações acima e de outras que forem importantes num manual (Conselho Federal de Psicologia, 2001, 2003).

A validade pode ser verificada sob diferentes perspectivas, enfocando-se o construto, o conteúdo e/ou o critério. Recorre-se às evidências de validade de critério quando se procura avaliar o grau com que o instrumento discrimina, de acordo com um critério padrão, pessoas que diferem em determinadas características. Quando o instrumento e o critério são aplicados simultaneamente, fala-se de validade concorrente; quando o critério é avaliado no futuro, fala-se de validade preditiva. Os resultados obtidos serão expressos em graus de concordância - identidade entre o teste e a outra forma -, denominada de relação entre eles. Ela é concebida como o grau de eficácia para predizer um determinado desempenho do sujeito, que deve ser medido por meio de técnicas independentes do próprio teste (Anastasi & Urbina, 2000; Pasquali, 1998).

Baseado na filosofia de que uma das medidas mais eficazes para melhorar a utilização dos testes é proporcionar uma boa informação e formação aos usuários, em 1995, o Colégio Oficial de Psicólogos da Espanha criou uma comissão de testes com o intuito de analisar os problemas envolvidos no seu uso. Em linhas gerais, o procedimento de avaliação proposto consiste em revisar os testes de forma sistemática por parte dos pesquisadores, utilizando um questionário inspirado nos modelos holandês e inglês atualmente em funcionamento. Um dos itens desse questionário diz respeito ao tamanho das amostras na validação do construto, que são consideradas: pequena ( $N < 200$ ); moderada ( $200 \leq N < 500$ ); grande ( $N \geq 500$ ). Especificado o tamanho da amostra, o item subsequente define a mediana das correlações do teste com outros testes similares de acordo com os seguintes critérios: inadequada ( $r < 0,25$ ); inadequada mas com algumas carências ( $0,25 \leq r < 0,40$ ); adequada ( $0,40 \leq r < 0,50$ ); boa ( $0,50 \leq r < 0,60$ ) e excelente ( $r \geq 0,60$ ) (Prieto & Muñiz, 2000).

Alguns pressupostos teóricos foram levados em consideração quanto ao uso dos instrumentos explorados neste trabalho. Em primeiro lugar, estão as contri-

buições da psicologia cognitiva e seus modelos explicativos sobre a aquisição e o desenvolvimento da leitura e escrita, tomando por base a teoria do processamento humano da informação. Esses modelos pressupõem a existência de objetos reais que podem ser medidos e modelados. Quando aplicada ao letramento, essa perspectiva sugere que os processos de leitura e escrita sejam subdivididos em subprocessos, cada um deles com uma função diferente (Irwin & Doyle, 1992). Em segundo lugar, está a tendência dos estudiosos a uma visão integradora da leitura e da escrita, entendidas como processos inter-relacionados, que não podem mais ser abordados separadamente (Kopke Filho, 2001; Pugh & Pawan, 1991; Santos, 1997). Considerou-se também que a leitura é uma atividade que constitui, ao mesmo tempo, forma de instrução e instrumento para o manejo de outras fases do currículo. Assim, a eficiência na leitura se relaciona de forma estreita com o êxito escolar (Alliende & Condemarín, 1987; Condemarín & Milicic, 1994).

Há que se ponderar, também, que a linguagem oral, a leitura e a escrita são consideradas instrumentais para a vida social e acadêmica e são o centro da maior parte das pesquisas em dificuldades de aprendizagem. A identificação dessas diferentes dificuldades tem ocorrido tardiamente, após vários fracassos, o que pode provocar desvios no comportamento da criança. No entanto, há como adquirir informações e desbloqueá-las, modificando todo o potencial dinâmico de aprendizagem, quando são recebidas intervenções psicopedagógicas adequadas. Para tanto é necessário que tais dificuldades sejam identificadas o mais precocemente possível, permitindo sua eventual superação (Jardim, 2001; Sisto, 2001; 2002).

Desse modo, instrumentos têm sido desenvolvidos para diagnosticar as dificuldades de aprendizagem referentes à linguagem. Duas das escalas aqui utilizadas - a de Avaliação de Dificuldades na Aprendizagem da Escrita (ADAPE) e a de Reconhecimento de Palavras - foram elaboradas por Sisto, respectivamente em 2001 e 2004. Ambas identificam a fase de alfabetização em que os alunos se encontram e conseguem detectar eventuais dificuldades de aprendizagem. O teste de Cloze (Santos, 2005) foi escolhido para avaliar a compreensão em leitura, podendo-se, por meio dele, acessar as inter-relações estabelecidas entre o leitor e o

texto escrito. Com esse instrumento também é possível identificar eventuais dificuldades de leitura (Kintsch & van Dijk, 1978; Condemarín & Milicic, 1994; Kintsch, 1994; Santos, Primi, Taxa & Vendramini, 2002).

Tomando por base as questões psicométricas mencionadas, o foco principal do presente trabalho será explorar evidências de validade convergente entre três instrumentos desenvolvidos para diagnosticar as dificuldades de aprendizagem de escrita e leitura em crianças. Para tanto, será considerada a análise da validação de construto de Campbell (1960), que salientou que, a fim de demonstrar a validade do construto, é necessário mostrar que um teste apresenta correlação alta com outras variáveis com as quais ele deve teoricamente correlacionar-se.

A literatura consultada não evidenciou a existência de estudos de validade convergente entre esses instrumentos, que supostamente avaliam construtos similares e envolvem as habilidades de leitura e de escrita. Para esse objetivo, serão consideradas as afirmações de Anastasi e Urbina (2000): a interpretação dos resultados obtidos pelos instrumentos, para que seja válida, depende de verificação para se avaliar se estão medindo aquilo que se propõem a medir. Outros tipos de validade também serão buscados. Primeiramente, serão averiguadas as habilidades linguísticas tal como mensuradas pelos três instrumentos, por série, para derivar evidências de validade de critério. Em seguida, será verificada a diferença entre as médias dos testes de Reconhecimento de Palavras e de Cloze pelas categorias do ADAPE também para derivar validade de critério. Finalmente serão correlacionados os escores dos instrumentos para buscar evidências de validade convergente.

## Método

### Participantes

Participaram deste estudo 266 crianças do ensino fundamental, entre 8 e 13 anos de idade ( $M=9,42$ ;  $DP=0,808$ ), de ambos os sexos: 140 (52,6%) do sexo masculino e 126 (47,4%) do feminino. Do total da amostra, 118 (44,4%) frequentavam a 3ª série e 148 (55,6%) a 4ª, sendo 66 (24,8%) de escolas particulares, 189 (71,1%) de públicas, quatro do SESI (1,5%) e sete não informaram (2,6%). Todas as escolas eram do interior do Estado de São Paulo.

## Instrumentos

*Questões de identificação:* os sujeitos informaram, na mesma folha do Reconhecimento de Palavras, os seguintes dados: nome, idade, sexo, série e instituição de ensino à qual pertenciam.

*Avaliação de Dificuldades na Aprendizagem da Escrita:* é composta por 114 palavras; 60 delas apresentam algum tipo de dificuldade, classificada como encontro consonantal, dígrafo, sílaba composta e sílaba complexa. Esse conjunto de palavras engloba as dificuldades mais comuns na escrita entre crianças de primeiras e segundas séries, identificadas por um levantamento das palavras comuns utilizadas no material didático dessas séries, segundo o autor Sisto (2001). Vale ressaltar que esse instrumento apresenta evidências de validade de critério e índices de precisão satisfatórios.

Os critérios de correção seguiram a escala de pontuação dos erros estabelecida pelo autor. Como não há critérios para a 4ª série, neste trabalho foram empregados os mesmos da 3ª série, a saber: até 10 erros - sem indícios de Dificuldade de Aprendizagem (DA); de 11 a 19 erros - DA leve; de 20 a 49 erros - DA média; 50 erros ou mais - DA acentuada.

*Reconhecimento de palavras:* contém 55 palavras organizadas em ordem alfabética. Para cada uma delas foram criadas outras duas com erros gramaticais. É um teste de múltipla escolha, com itens que têm três alternativas de resposta, das quais somente uma é correta. No estudo realizado por Sisto (2004), a análise da variância entre as médias dos participantes indicou uma diferença estatisticamente significativa entre as séries ( $p<0,001$ ), verificada pelo teste de Tukey, que separou os escores nos grupos correspondentes às séries frequentadas, atribuindo a validade de critério ao instrumento.

*Teste de Cloze:* foi utilizado um texto estruturado na forma do Cloze tradicional com os quintos vocábulos omitidos, substituindo-os por um traço proporcional ao tamanho da palavra omitida, que deveria ser recuperada pelo leitor para restituir o sentido completo à seleção, tal como preconizado por Taylor (1953). Esse texto foi especialmente montado para ser utilizado com crianças da faixa etária do ensino fundamental. Elaborado por Santos (2005), também apresenta evidências de validade.

## Critérios de correção

Foram utilizados os critérios estabelecidos pelos autores dos instrumentos. Para o ADAPE foi atribuído um ponto para cada palavra grafada erradamente pela criança, pelas ausências de palavras, acentuação errada e uso indevido de maiúsculas e minúsculas. A pontuação de cada criança corresponde à soma dos erros cometidos, sendo o escore máximo de 114 pontos.

No Reconhecimento de Palavras, a cada acerto foi atribuído um ponto, sendo possível atingir o escore máximo de 55 pontos. Também no Teste de Cloze foi atribuído um ponto para cada acerto, sendo consideradas corretas as palavras que coincidirem com as usadas pelo autor do texto. Essa forma de correção é denominada literal ou *verbatim* e é recomendada por evitar o subjetivismo na avaliação. Assim, o número de acertos máximos é 15, correspondentes ao número de omissões.

## Procedimentos

Esta pesquisa foi analisada e aprovada pelo Comitê de Ética e Pesquisa da Universidade São Francisco (protocolo 209/02). A coleta de dados ocorreu em instituições cujas diretoras autorizaram sua realização e com aqueles alunos cujos pais assinaram o Termo de Consentimento Esclarecido, que permite a participação na pesquisa.

Os instrumentos foram aplicados nos alunos, pela professora, em situação de sala de aula. Para evitar o efeito de fadiga, cada instrumento foi aplicado em diferentes momentos de um dia normal de atividades escolares, entre outras atividades previstas no planejamento da aula, na seguinte sequência: ADAPE, Re-

conhecimento de Palavras e Teste de Cloze. Também foi fornecido um roteiro às professoras com instruções para aplicação de cada instrumento.

## Resultados

Buscou-se explorar eventuais diferenças nos escores dos instrumentos em razão da série frequentada. As estatísticas descritivas e o resultado do teste *t* de Student para a comparação entre as séries estão apresentados na Tabela 1.

Verificou-se, na comparação entre as séries, que os escores dos três instrumentos foram melhores para a 4ª série ( $n=148$ ) em relação aos da 3ª série ( $n=118$ ). As diferenças entre as médias por série revelaram um índice estatisticamente significativo somente para o Reconhecimento de Palavras ( $t=-2,050$ ;  $p=0,041$ ), conferindo-lhe a validade de critério. No teste de Cloze a diferença foi marginalmente significativa ( $t=-1,924$ ;  $p=0,055$ ).

Para averiguar se havia diferença significativa entre as médias do Reconhecimento de Palavras e Cloze pelas categorias do ADAPE, procedeu-se à análise de variância, que indicou, respectivamente, as seguintes diferenças estatisticamente significativas:  $[F(3,262) = 105,942$ ;  $p < 0,001]$  e  $[F(3,262) = 49,734$ ;  $p < 0,001]$ .

Para verificar quais categorias justificavam as diferenças encontradas, utilizou-se o teste *post-hoc* de Tukey, das categorias do ADAPE em relação aos escores do Cloze. Os quatro agrupamentos resultantes aparecem na Tabela 2.

O teste de Tukey separou os escores do Reconhecimento de Palavras nas quatro categorias do ADAPE, constatando a diferença estatisticamente significativa entre os escores dos instrumentos. Pode-se inferir, a partir

**Tabela 1.** Estatísticas descritivas e teste *t* de Student para comparação entre as séries. Bragança Paulista (SP), 2005.

Escolaridade		n	Média	Desvio-padrão	t	p
ADAPE	3	118	31,34	23,62	0,695	0,487
	4	148	29,39	22,06		
Reconhecimento de palavras	3	118	38,48	10,85	-2,050	0,041
	4	148	41,09	9,83		
CLOZE	3	118	9,41	3,86	-1,924	0,055
	4	148	10,30	3,70		

ADAPE: escala de avaliação de dificuldades na aprendizagem da escrita.

desses resultados, que o Reconhecimento de Palavras é sensível para captar diferenças tal como separadas pelas categorias do ADAPE, concedendo-lhe também essa validade de critério.

Realizou-se também o teste de *post-hoc* Tukey para identificar quais categorias do ADAPE se diferenciaram estatisticamente em relação aos escores do teste de Cloze. Os agrupamentos dos resultados do teste de Cloze são mostrados na Tabela 3.

Foram formados três grupos de categorias do ADAPE de acordo com os escores do Cloze. O teste de Tukey agrupou as categorias "DA leve" e "Sem indícios de DA", cujas médias foram muito parecidas, respecti-

**Tabela 2.** Distribuição das categorias do ADAPE em relação aos escores do Reconhecimento de Palavras. Bragança Paulista (SP), 2005.

ADAPE por categorias	n	Média no reconhecimento de palavras			
		1	2	3	4
DA acentuada	42	26,38			
DA média	128		37,94		
DA leve	61			47,21	
Sem indícios de DA	35				50,80

ADAPE: escala de avaliação de dificuldades na aprendizagem da escrita; DA: dificuldade de aprendizagem.

**Tabela 3.** Distribuição das categorias do ADAPE em relação aos escores do teste de Cloze. Bragança Paulista (SP), 2005.

ADAPE por categorias	n	Média no Cloze		
		1	2	3
DA acentuada	42	5,48		
DA média	128		9,60	
DA leve	61			12,02
Sem indícios de DA	35			12,69

ADAPE: escala de avaliação de dificuldades na aprendizagem da escrita; DA: dificuldade de aprendizagem.

**Tabela 4.** Índices de correlações (*r*) e níveis de significâncias (*p*) entre os escores dos instrumentos. Bragança Paulista (SP), 2005.

		Reconhecimento de palavras	Cloze
Cloze	<i>r</i>	0,651	-
	<i>p</i>	<0,001	-
ADAPE	<i>r</i>	-0,760	-0,605
	<i>p</i>	<0,001	<0,001

ADAPE: escala de avaliação de dificuldades na aprendizagem da escrita.

vamente 12,02 e 12,69. Esses resultados concedem ao teste de Cloze validade de critério por diferenciar de maneira similar três das quatro categorias do ADAPE.

Em conformidade com o objetivo principal deste estudo, buscou-se a validade convergente entre os instrumentos. Para tanto, foram realizadas provas de correlação de *Pearson* com os escores alcançados no ADAPE, no Cloze, e no Reconhecimento de Palavras (Tabela 4).

Entre os acertos do Cloze e do Reconhecimento de Palavras obteve-se um índice de correlação positivo e significativo ( $p < 0,001$ ). Entre os erros do ADAPE e os acertos do Reconhecimento de Palavras, o índice de correlação foi negativo e significativo ( $p < 0,001$ ), assim como entre os erros do ADAPE e os acertos do Cloze ( $p < 0,001$ ). Dessa maneira, ficou mais uma vez evidenciada a correlação entre os escores médios dos três instrumentos analisados.

## Discussão

Com a preocupação de utilizar instrumentos de avaliação cujas propriedades psicométricas sejam estabelecidas para populações específicas, foram exploradas, neste estudo, evidências de validade convergente entre três instrumentos desenvolvidos para diagnosticar as dificuldades de aprendizagem de escrita e leitura em crianças. Para esse objetivo, foi considerado que a interpretação dos resultados obtidos pelos instrumentos, para que seja válida, depende de verificação para avaliar se os instrumentos estão medindo aquilo que se propõem a medir.

Desse modo, primeiramente, foram averiguadas, por série, as habilidades linguísticas tal como mensuradas pelos três instrumentos para derivar evidências de validade de critério. Em segundo lugar, foi verificada a diferença entre as médias dos testes de Reconhecimento de Palavras e de Cloze pelas categorias do ADAPE também para derivar validade de critério. Finalmente foram correlacionados os escores dos instrumentos para buscar a validade convergente.

Na comparação entre as séries, os escores dos três instrumentos foram melhores para a 4ª série ( $n=148$ ) do que para a 3ª série ( $n=118$ ). Esse resultado era esperado, pois a automatização de processos permite que

quanto mais os subprocessos vão se tornando automáticos, mais os leitores e escritores vão se tornando hábeis para atender à demanda de atividades cognitivas (Irwin & Doyle, 1992). Verificou-se com esses resultados que há evidências de validade de critério concorrente em relação à sensibilidade dos instrumentos para captar a evolução do desenvolvimento das crianças na progressão das séries (Anastasi & Urbina, 2000; Pasquali, 1998). No entanto, as diferenças entre as médias revelaram um índice estatisticamente significativo somente para o Reconhecimento de Palavras.

Na análise da variância entre as médias do Reconhecimento de Palavras e Cloze, pelas categorias do ADAPE, foram encontradas diferenças estatisticamente significativas para ambos os instrumentos, indicando que eles avaliam construtos similares. Para identificar quais categorias se diferenciaram estatisticamente, foi realizado o teste de Tukey tanto para o Reconhecimento de Palavras como para o Cloze.

O teste de Tukey separou os escores do Reconhecimento de Palavras nas quatro categorias do ADAPE, confirmando a diferença estatisticamente significativa entre os escores dos instrumentos. Pode-se inferir, assim, que com o Reconhecimento de Palavras se obtém a mesma classificação quanto às habilidades de compreensão de leitura que o ADAPE nas habilidades de escrita das crianças.

Esse resultado confere ao Reconhecimento de Palavras a mesma validade de critério do ADAPE, visto que ele consegue obter as diferenças na compreensão de leitura nos mesmos níveis de alfabetização. Ficou demonstrado, dessa maneira, que o Reconhecimento de Palavras discriminou características que diferem as crianças de acordo com um critério padrão, que foi aplicado simultaneamente com o instrumento, atribuindo-se assim a validade de critério concorrente (Anastasi & Urbina, 2000; Pasquali, 1998).

De acordo com os escores do Cloze, o teste de Tukey formou três grupos de categorias do ADAPE, porque as categorias "DA leve" e "Sem indícios de DA" se agruparam. É importante destacar que o que pode ter influenciado esse agrupamento foi o fato de o número de alunos dessas categorias ser bastante desigual, respectivamente 61 e 35. Mas, mesmo assim, os instru-

mentos conseguiram medir as habilidades em faixas muito próximas. Infere-se daí que também se pode atribuir a validade de critério ao Cloze pelos níveis de alfabetização do ADAPE, pois ficou demonstrado que houve discriminação de características que diferem as crianças de acordo com um critério padrão. Como o critério foi aplicado junto com o instrumento, atribuiu-se a validade de critério concorrente (Anastasi & Urbina, 2000; Pasquali, 1998).

Esses últimos resultados, que classificaram as crianças em faixas muito próximas nos três instrumentos, confirmam a relação entre as habilidades de leitura e escrita, pois, atualmente, há uma visão integradora da leitura e da escrita como processos inter-relacionados que não podem mais ser abordados separadamente e são vistos como resultantes da ação do pensamento integrado, simultaneamente dinâmico e complexo. Assim sendo, os escores obtidos por meio dos instrumentos utilizados demonstraram que o desempenho em leitura pode estar relacionado à facilidade para o reconhecimento de palavras, à melhor utilização da informação gráfica - com uso dos conhecimentos fonológico, ortográfico e do vocabulário visual, adquiridos com a prática de leitura - e à utilização eficiente das informações contextuais (Irwin & Doyle, 1992; Kopke Filho, 2001; Pugh & Pawan, 1991; Santos, 1997).

Foram então realizadas provas de correlação de *Pearson* com os resultados obtidos nos três instrumentos. Obteve-se um índice de correlação positivo e significativo ( $r=0,651$ ) entre os acertos do Cloze e do Reconhecimento de Palavras. Esse índice é classificado como excelente ( $r\geq 0,60$ ) de acordo com o critério "Mediana das correlações do teste com outros testes similares", do *Cuestionario para evaluación de los tests* de Prieto e Muñiz (2000).

De maneira similar, os erros do ADAPE e os acertos do Reconhecimento de Palavras tiveram um índice de correlação significativo ( $r=-0,760$ ), porém negativo, também considerado excelente ( $r\geq 0,60$ ) pelos critérios expostos anteriormente. Pode-se inferir, então, que os problemas na escrita independeram do tipo de estímulo utilizado: o auditivo no caso do ADAPE e o auditivo e visual no caso do Reconhecimento de Palavras. Deduz-se, então, que esses construtos são similares e dependem de processos cognitivos imbricados entre si.

Entre os erros no ADAPE e os acertos no Cloze houve um índice negativo e significativo ( $r=-0,605$ ), que é considerado excelente de acordo com os parâmetros de Prieto e Muñiz (2000). Mais uma vez ficou demonstrado que quanto mais as crianças erraram na escrita, menos compreensão tiveram na leitura.

Em resumo, foi demonstrado pelo levantamento bibliográfico que o desenvolvimento dos testes ADAPE, Reconhecimento de Palavras e Cloze emergiu da literatura científica, ficando definido o construto e avaliada sua correta representação nos testes. Além disso, os instrumentos foram submetidos a processos de desenvolvimento e validação para assegurar que realizam o objetivo a que se destinam. Assim, ficaram esclarecidas as naturezas dos traços sob consideração e as condições que afetam seu desenvolvimento e suas manifestações, conforme preconizaram os autores levantados na revisão bibliográfica. Unindo os resultados alcançados a esses preceitos psicométricos, conclui-se que os três instrumentos avaliam construtos semelhantes e apresentam correlação alta com as variáveis com as quais eles devem se correlacionar. Dessa maneira, infere-se que esses instrumentos medem o que se propõem a medir e que há evidências de validade convergente entre eles (Anastasi & Urbina, 2000; Campbell, 1960; Conselho Federal de Psicologia, 2001, 2003; Cronbach & Meehl, 1955; Muñiz, 2004; Pasquali, 2003; Ramsay & Reynolds, 2000).

Faz-se necessário, contudo, dar continuidade à construção e à busca de evidências de validade dos instrumentos psicológicos no Brasil a fim de solucionar sua precariedade e seu suprimento com a utilização de instrumentos estrangeiros (Noronha, 2002; Noronha et al., 2001; Noronha et al., 2003; Noronha et al., 2004; Sisto, Codenotti, Costa & Nascimento, 1979; Vendramini & Noronha, 2002).

## Conclusão

Para entender as razões que levam uma criança a executar uma tarefa cognitiva de forma pior que a criança normal é necessário saber como seria a resolução correta da tarefa e utilizar esse conhecimento para analisar onde residem os problemas para a criança com dificuldade de aprendizagem. Na escola, as dificuldades

que interessam são aquelas que impedem o êxito acadêmico e, nesse sentido, os instrumentos de medida das habilidades de escrita e de leitura desempenham um papel importante no diagnóstico, pois a identificação das dificuldades possibilita analisar melhor a intervenção a ser proposta. Há que se levar em conta, principalmente, que as crianças com dificuldades de aprendizagem podem modificar seu potencial de aprendizagem ao adquirirem informações que as desbloqueiem.

Como foi sinalizado, o Brasil carece de instrumentos para medir os níveis de alfabetização, de forma a expressar a qualidade do ensino no país. As políticas públicas preocupam-se com as estatísticas; no caso da educação, com os números de ingressantes e egressos. Em relação a esse assunto, há uma vigorosa discussão sobre os efeitos da progressão continuada para o favorecimento dessas estatísticas.

Conforme apresentado neste trabalho, a psicologia cognitiva vem permitindo avanços significativos tanto no conhecimento da aquisição da linguagem, como no das dificuldades de aprendizagem. Porém, constata-se que os Parâmetros Curriculares Nacionais não levam em conta, ignoram, ou minimizam a literatura científica.

As implicações dessa atitude se refletem nos indicadores do Sistema de Avaliação do Ensino Básico (SAEB), programa de avaliação implementado pelo governo brasileiro. Ainda que esse programa seja pautado por uma dimensão humana com uma longa tradição de pesquisa na psicologia, a das habilidades cognitivas, seus métodos de mensuração têm sido muito criticados, de acordo com os argumentos já expostos.

Considerando a importância dos aspectos apresentados em relação às dificuldades de aprendizagem e a necessidade de avaliação precoce das crianças em fase escolar, de forma a prevenir, e não apenas remediar o surgimento de dificuldades de aprendizagem associadas ao processo de escolarização, compreende-se que os testes são ferramentas úteis e necessárias. No entanto, no Brasil, eles caíram no descrédito por parte dos usuários por um longo tempo. Essa situação foi gerada pela inexistência de testes nacionais, o que levou



ao uso dos estrangeiros, adaptados ou não, mas mesmo assim sem supervisão de um órgão oficial, o que é, atualmente, inconcebível, visto que, para serem considerados válidos, os testes passam por rigoroso tratamento científico, como visto neste trabalho. Cabe retomar que, mesmo que os testes contribuam de forma valiosa, oferecendo informações detalhadas a respeito do desempenho da criança, o seu uso merece cautela, pois eles não podem ser usados como única medida: deve-se considerar a existência de vieses culturais e o contexto em que os instrumentos são aplicados.

Vale destacar, entretanto, que embora esses instrumentos sejam propostas inovadoras e promissoras de avaliação, é preciso pesquisar mais evidências de validade para propiciar a ampliação do conhecimento sobre suas características psicométricas. Espera-se ainda a realização de outras pesquisas que superem as limitações deste estudo, principalmente no que se refere ao tamanho da amostra, e que possam contribuir para a ampliação dos conhecimentos na área e para aprimorar os procedimentos técnicos empregados na avaliação psicológica e os próprios instrumentos.

## Referências

- Alliende, F., & Condemarín, M. (1987). *Leitura: teoria, avaliação e desenvolvimento*. Porto Alegre: Artes Médicas.
- American Educational Research Association. (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Anastasi, A., & Urbina, S. (2000). *Testagem psicológica*. Porto Alegre: Artes Médicas.
- Campbell, D. T. (1960). Recommendations for APA test standards regarding construct, trait, and discriminant validity. *American Psychologist*, 15 (4), 546-553.
- Condemarín, M., & Milicic, N. (1994). *Test de Cloze: procedimiento par el desarrollo y la evaluación de la comprensión lectora*. Santiago de Chile: Andrés Bello.
- Conselho Federal de Psicologia. (2001). *Resolução nº 025/2001*. Recuperado em agosto 21, 2006, disponível em <http://www.pol.org.br>
- Conselho Federal de Psicologia. (2003). *Resolução nº 002/2003*. Recuperado em agosto 21, 2006, disponível em <http://www.pol.org.br>
- Cronbach, L. J. (1960). *Essentials of psychological testing* (2<sup>nd</sup> ed.). New York: Harper & Brothers.
- Cronbach, L. J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52 (2), 281-302.
- Goldstein, G., & Hersen, M. (2000). Historical perspectives. In G. Goldstein & M. Hersen (Orgs.), *Handbook of psychological assessment* (pp.3-17). Kidlington, Oxford: Pergamon.
- Irwin, J.W., & Doyle, M. A. (1992). *Reading/writing connections: learning from research*. Newark: IRA.
- Jardim, W. R. S. (2001). *Dificuldades de aprendizagem no ensino fundamental: manual de identificação e intervenção*. São Paulo: Edições Loyola.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85 (5), 363-394.
- Kintsch, W. (1994). Text comprehension, memory and learning. *American Psychologist*, 49 (4), 294-303.
- Kopke Filho, H. (2001). *Estratégias em compreensão da leitura: conhecimento e uso por professores de língua portuguesa*. Tese de doutorado não-publicado, Universidade de São Paulo.
- Muñiz, J. (2004). La validación de los testes. *Metodología de las Ciencias del Comportamiento*, 5 (2), 121-141.
- Noronha, A. P. P. (2002). Problemas mais graves e mais frequentes no uso dos testes psicológicos. *Psicologia Reflexão e Crítica*, 15 (1), 135-142.
- Noronha, A. P. P., Freitas, F. A., & Ottati, F. (2001). Informações contidas nos manuais de testes de inteligência. *Psicologia em Estudo*, 6 (2), 101-106.
- Noronha, A. P. P., Freitas, J. V., Baldo, C. R., Barbini, P. F., & Almeida, M. C. (2004). Conhecimento de estudantes a respeito de conceitos de avaliação psicológica. *Psicologia em Estudo*, 9 (2), 263-269.
- Noronha, A. P. P., Oliveira, K. L., & Beraldo, F. N. M. (2003). Instrumentos psicológicos mais conhecidos e utilizados por estudantes e profissionais de Psicologia. *Psicologia Escolar e Educacional*, 7 (1), 47-56.
- Pasquali, L. (1998). Princípios de elaboração de escalas psicológicas. *Revista de Psiquiatria Clínica*, 25 (5), 206-213.
- Pasquali, L. (2003). *Psicometria: teoria dos testes na psicologia e na educação*. Petrópolis: Vozes.
- Prieto, G., & Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77 (1), 65-75.
- Pugh, S. L., & Pawan, F. (1991). Reading, Writing and Academic Literacy. In R. F. Flippo & D. C. Caverly (Eds.), *College reading and study strategy programs*. Newark, Delaware: IRA.
- Ramsay, M. C., & Reynolds, C. R. (2000). Development of a scientific test: a practical guide. In G. Goldstein & M. Hersen (Orgs.), *Handbook of psychological assessment* (pp.21-42). Kidlington, Oxford: Pergamon.
- Santos, A. A. A. (1997). Programas de remediação: uma alternativa para o desenvolvimento da compreensão em leitores adultos. *Pro-posições*, 5 (1), p.115.
- Santos, A. A. A. (2005). Evidências de validade de critério para o teste de Cloze. Manuscrito não-publicado, Pro-

- grama de Pós-Graduação *Stricto-Sensu* em Psicologia, Universidade São Francisco, Itatiba.
- Santos, A. A. A., Primi, R., Taxa, F., & Vendramini, C. M. M. (2002). O teste de Cloze na avaliação da compreensão em leitura. *Psicologia: Reflexão e Crítica*, 15 (3), 549-560.
- Sisto, F. F. (2001). Dificuldade de aprendizagem em escrita: um instrumento de avaliação (ADAPE). In F. F. Sisto, E. Boruchovitch & L. D. T. Fini (Orgs.), *Dificuldades de aprendizagem no contexto psicopedagógico* (pp.190-213). Petrópolis: Vozes.
- Sisto, F. F. (2002). Avaliação de dificuldade de aprendizagem: uma questão em aberto. In F. F. Sisto, E. A. Dobránszky & A. Monteiro (Orgs.), *Cotidiano escolar: questões de leitura, matemática e aprendizagem* (pp.121-141). Petrópolis: Vozes.
- Sisto, F. F. (2004). Evidências de validade de critério para a escala de reconhecimento de palavras. Manuscrito não-publicado, Programa de Pós-graduação *Stricto-Sensu* em Psicologia, Universidade São Francisco, Itatiba.
- Sisto, F. F., Codenotti, N., Costa, C. A. J., & Nascimento, T. C. N. (1979). Testes psicológicos no Brasil: o que medem realmente. *Educação e Sociedade*, 2 (2), 152-165.
- Spies, R. A., & Plake, B. S. (2005). *The Sixteenth mental measurements yearbook*. Nebraska: Harbound.
- Taylor, W. L. (1953). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly*, 30 (3), 415-433.
- Vendramini, C. M. M., & Noronha, A. P. P. (2002). Estudo comparativo entre testes de inteligência e de personalidade. *Psico*, 33 (2), 413-426.

Recebido em: 14/4/2008

Versão final reapresentada em: 26/2/2010

Aprovado em: 4/3/2010