

Laura de Carvalho



**EVIDÊNCIAS DE VALIDADE DO SISTEMA DE
PONTUAÇÃO GRADUAL DO BENDER (B-SPG)**

**Itatiba
2006**

Laura de Carvalho

**EVIDÊNCIAS DE VALIDADE DO SISTEMA DE
PONTUAÇÃO GRADUAL DO BENDER (B-SPG)**

Tese de doutorado apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Psicologia da Universidade São Francisco para obtenção do título de Doutor em Psicologia.

ORIENTADORA: Prof^a. Dr^a. Ana Paula Porto Noronha

Itatiba
2006

Universidade São Francisco

PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU*

DOUTORADO EM PSICOLOGIA

EVIDÊNCIAS DE VALIDADE DO SISTEMA DE PONTUAÇÃO GRADUAL DO BENDER (B-SPG)

Autora: Laura de Carvalho

Orientadora: Prof^a.Dr^a. Ana Paula Porto Noronha

Este exemplar corresponde à redação final da tese de doutorado de Laura de Carvalho.

Data: ____ / ____ / ____

COMISSÃO EXAMINADORA

Prof^a. Dr^a. Ana Paula Porto Noronha (presidente)

Prof^a. Dr^a. Selma de Cássia Martinelli

Prof^a Dr^a. Evely Boruchovitch

Prof^a. Dr^a. Acácia Aparecida Angeli dos Santos

Prof^a. Dr^a. Maria Cristina Rodrigues Azevedo Joly

**Itatiba
2006**

Dedicatória

Ao meu companheiro, Pedro Mazzaro, e aos nossos filhos, Eduardo e Gustavo, especialmente pelo apoio e incentivo, e porque muitas vezes os deixei sozinhos, para que pudesse realizar as pesquisas. Este trabalho só foi possível porque somaram-se esforços conjuntos.

Agradecimentos

À minha orientadora, Profa. Dra. Ana Paula Porto Noronha, que confiou em mim desde o primeiro contato já no processo seletivo, e depois como orientadora muito me ensinou, apoiou, e incentivou quando foi necessário. Obrigada!

À minha família, pelo apoio e carinho que me demonstrou a cada dia. Pela força e por estar sempre do meu lado ao longo dessa caminhada. Sem dúvida nenhuma, é a mola propulsora da minha existência.

À minha mãe, Iracema Rodrigues de Carvalho (*in memoriam*) que sempre me ajudou com palavras meigas e carinhosas, apoiando-me e valorizando-me em todas as iniciativas de buscar o conhecimento. Sem dúvida, ela foi uma vencedora.

Aos meus colegas de turma, que muito me incentivaram e contribuíram para a construção do meu conhecimento e crescimento profissional. E pelas divertidas “piadas” no momento do cafezinho.

A todos os Professores do Programa de Pós-Graduação em Psicologia da Universidade São Francisco, especialmente aos Professores Fermino Fernandes Sisto e Acácia Aparecida Angeli dos Santos, pelas valiosas contribuições que fizeram ao trabalho.

Às professoras Dras. Evely Boruchovich, Maria Cristina Rodrigues Azevedo Joly e Selma de Cássia Martinelli pelas contribuições feitas na ocasião da qualificação. À equipe da Secretaria, em especial à colega Roseli, pela dedicação e carinho com que sempre me atendia.

Às estagiárias do Curso de Graduação em Psicologia da Unoeste de Presidente Prudente/SP, Tânia Michiuti e Luciana, pela colaboração na coleta dos dados.

Aos colegas Gleiber, Fabiano, Monalisa, Neide, Daniel e Márcia pelas dicas na revisão do material e a todos que, direta ou indiretamente, participaram dessa caminhada, o meu carinho e agradecimento.

Resumo

Carvalho, L. (2006). *Evidências de Validade do Sistema de Pontuação Gradual do Bender (Bender-SPG)*. Tese de Doutorado, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba, 146 p.

Este trabalho teve por objetivo investigar evidências de validade do Teste Gestáltico Visomotor de Bender - Sistema de Pontuação Gradual (B-SPG), utilizando para tanto os instrumentos Teste de Compreensão em leitura (Cloze), Escala de Avaliação de Dificuldade na Aprendizagem da Escrita (ADAPE) e o Teste de Reconhecimento de Palavras. Partiu-se da premissa de que os testes mencionados apresentariam em comum certos aspectos como o fato de serem utilizados para predição das capacidades de leitura e escrita de crianças e envolverem aspectos motores (evidências de validade de construto). Procuraram-se, assim, relações entre a aprendizagem e a maturidade visomotora de crianças (evidência de validade concorrente). Os escores foram relacionados, separadamente, por sexo, idade, série e tipos de escola. Além disso, os grupos extremos formados em razão dos instrumentos foram analisados em relação a possíveis diferenças nas pontuações do Bender. Participaram desse estudo 297 crianças de ambos os sexos, com idades entre 7 e 10 anos, e a média entre as idades dos sujeitos era de 9,04 (DP=0,88), sendo 147 do sexo feminino (48,5%) e 156 do sexo masculino (51,5%). Quanto à escolaridade, 18 pertenciam a 1ª série (6,1%); 24, a 2ª série (8,1%); 118 freqüentavam a 3ª série (39,7%) e 137 estavam na 4ª série (46,5%). Desses, 81 eram de Escola Particular (27,3%) e 216, de Escola Pública (72,7%) e cursavam de primeira a quarta série do ensino fundamental no interior do estado de São Paulo. Aplicaram-se os testes coletivamente, sendo que no caso do Bender os *slides* das Figuras foram projetados na parede para que pudessem ser copiadas pelas crianças. Foram evidenciadas correlações significativas entre os escores dos testes e o Bender-SPG, bem como diferenças entre os grupos extremos dos instrumentos em relação ao escore do Teste de Bender em todas as situações estudadas. Concluiu-se que o Teste Gestáltico de Bender-SPG mostrou-se sensível para captar as diferenças entre crianças com e sem dificuldade de aprendizagem, bem como as diferenças relacionadas às séries, sexo, idade e instituições pesquisadas.

Palavras-chave: Desenvolvimento Visomotor, Teste de Bender, Avaliação Psicológica, Aprendizagem.

Abstract

Carvalho, L. (2006). Validity Evidences of the Bender System of Gradual Punctuation (B-SPG). Thesis of Doctorate, Program of *Stricto Sensu* Post graduation in Psychology, University of São Francisco, Itatiba, 146 p.

This work aims to investigate validity evidences of the Gestaltic Visomotor Bender - System of Gradual Punctuation (B-SPG) Test, using the following instruments: Cloze test, Evaluation of Difficulty in the Writing Learning Scale (ADAPE) and the Words Recognition Test. We postulate the premise that these tests would have certain aspects in common, such as the fact of being used for the prediction of children reading and writing abilities and involving motor aspects (evidences of construct validity). We searched relationship between learning and children maturity (evidence of competitive validity). The scores were separately related in terms of sex, age, school-level and types of school. Besides this, the extreme groups, which were constituted due to the instruments, were analysed in relation to possible differences in Bender punctuations. Participate in these investigation 297 children's, of both sexes and aged from seven to ten years-old. The median between the children ages was 9,04 (DP = 0,88), considering 144 girls (48,5%) and 153 boys (51,5%). In terms of education, 18 children attended the 1^a level, (6,1%); 24, second level (8,1%), 118 attended the third level (39,7%) and 137 attended the fourth one (46,5%). Among them, 81 studied at private schools (27,3%), 216 at public schools (72,7%) and attended the elementary education in school in the interior of São Paulo state. We applied the tests collectively, considering that the Bender slides were projected through the wall in order to be copied by the children. Significant correlations were evidenced between the scores tests and the Bender-SPG, as well as differences among the extreme groups of the instruments in relation to the Bender test score in all of the analysed situations. We conclude that the Gestaltic Bender Test (SPG) seemed to be sensible to capture the differences between children with and without learning school-levels, sex, age and the researched institutions.

Keywords: Development Visomotor; Bender Test; Psychological Evaluation; Assessment.

SUMÁRIO

LISTA DE FIGURAS.....	xi
LISTA DE TABELAS	xii
LISTA DE ANEXOS.....	xiv
APRESENTAÇÃO	1
CAPÍTULO I.....	8
TESTE BENDER GESTÁLTICO: HISTÓRICO E PRESSUPOSTOS TEÓRICOS.....	8
PRINCIPAIS SISTEMAS DE CORREÇÃO E ATRIBUIÇÕES DO TESTE DE BENDER.....	12
ESTUDOS RELACIONADOS AOS QUESTIONAMENTOS SOBRE OS CRITÉRIOS DE AVALIAÇÃO.....	24
CAPÍTULO II	31
PARÂMETROS PSICOMÉTRICOS: PRECISÃO, VALIDADE E PADRONIZAÇÃO DOS TESTES PSICOLÓGICOS	31
PESQUISAS COM O BENDER-SPG	39
OBJETIVOS	44
CAPÍTULO III	45
MÉTODO.....	45
Participantes	45
Instrumentos	45
O TESTE GESTÁLTICO VISOMOTOR DE BENDER: SISTEMA DE PONTUAÇÃO GRADUAL (B-SPG)	46
O TESTE CLOZE NA AVALIAÇÃO DA COMPREENSÃO EM LEITURA.....	48
ESCALA DE AVALIAÇÃO DE DIFICULDADE NA APRENDIZAGEM DA ESCRITA (ADAPE)	53

TESTE DO RECONHECIMENTO DE PALAVRAS	60
Procedimento	61
CAPÍTULO IV	65
Resultados e Discussão	65
CAPÍTULO V	104
Considerações Finais	104
VI REFERÊNCIAS.....	108
VII ANEXOS.....	121

LISTA DE FIGURAS

Figura 1 - Resultado da Média por Série no Teste Bender-SPG	79
Figura 2 - Resultado da Média por Série no Teste Reconhecimento de Palavras	80
Figura 3 - Resultado da Média por Série no Teste de Compreensão em Leitura (Cloze)	81
Figura 4 - Resultado da Média por Série no Teste ADAPE	82
Figura 5 - Resultado da Média por Idade no Teste Bender-SPG	87
Figura 6 - Resultado da Média por Idade no Teste Reconhecimento de Palavras	88
Figura 7 - Resultado da Média por Idade no Teste de Compreensão em Leitura (Cloze)	89
Figura 8 - Resultado da Média por Idade no Teste ADAPE	90
Figura 9 - Resultado da Média por Escola no Teste ADAPE	94
Figura 10 - Resultado da Média por Escola no Teste Bender-SPG	95
Figura 11 - Resultado da Média por Escola no Teste de Compreensão em Leitura (Cloze)	96
Figura 12 - Resultado da Média por Escola no Teste Reconhecimento de Palavras	97

LISTA DE TABELAS

Tabela 1 - Estatística descritiva dos instrumentos por série	66
Tabela 2 - Estatística descritiva dos instrumentos por tipo de escola	67
Tabela 3 - Estatística descritiva dos instrumentos por idade	68
Tabela 4 - Correlação de <i>Pearson</i> entre o Teste Bender-SPG e os instrumentos Reconhecimento de Palavras, Cloze e ADAPE	69
Tabela 5 - Correlação de <i>Pearson</i> entre os Testes Bender, Reconhecimento de Palavras, Cloze e ADAPE e as variáveis Idade e Escolaridade	72
Tabela 6 - Grupos extremos do Bender em comparação com o Teste Reconhecimento de Palavras	73
Tabela 7 - Grupos extremos do Bender em comparação com o Teste de Compreensão em Leitura (Cloze)	73
Tabela 8 - Grupos extremos do Bender em comparação com o Teste ADAPE	74
Tabela 9 - Índices de Precisão do Bender-SPG por Alfa de Cronbach	75
Tabela 10 - Média, desvio-padrão, <i>t</i> e <i>p</i> dos participantes divididos por gênero para todos os instrumentos	76
Tabela 11 - Comparação entre as médias de crianças de diferentes séries com base nos instrumentos por meio da ANOVA	78
Tabela 12 - Subconjunto formado pela prova de <i>Tukey</i> em relação à pontuação total do Teste de Bender-SPG para a série	83
Tabela 13 - Subconjuntos formados pela prova de <i>Tukey</i> em relação à pontuação total do Teste Reconhecimento de Palavras para a série	84

Tabela 14 - Subconjuntos formados pela prova de <i>Tukey</i> em relação à pontuação total do Teste de Compreensão em Leitura (Cloze) para a série	84
Tabela 15 - Subconjuntos formados pela prova de <i>Tukey</i> em relação à pontuação total do Teste ADAPE para a série	85
Tabela 16 - Comparação entre as médias de crianças de diferentes idades com base nos instrumentos por meio da ANOVA	86
Tabela 17 - Subconjuntos formados pela prova de <i>Tukey</i> em relação à pontuação no Bender para a idade	91
Tabela 18 - Subconjuntos formados pela prova de <i>Tukey</i> em relação à pontuação no teste Reconhecimento de Palavras para a idade	92
Tabela 19 - Subconjuntos formados pela prova de <i>Tukey</i> em relação à pontuação no Teste de Compreensão em Leitura (Cloze) para a idade	92
Tabela 20 - Subconjuntos formados pela prova de <i>Tukey</i> em relação à pontuação no ADAPE para a idade	93

LISTA DE ANEXOS

Anexo 1 - Teste Gestáltico Visomotor de Bender	122
Anexo 2 - Teste de Compreensão em Leitura (Cloze)	124
Anexo 3 - Teste da Escala de Avaliação de Dificuldades na Aprendizagem da Escrita (ADAPE)	125
Anexo 4 - Teste do Reconhecimento de Palavras	127
Anexo 5 - Termo de Consentimento Livre e Esclarecido	132

APRESENTAÇÃO

Os psicólogos estão, cada vez mais, sendo requisitados pela comunidade, órgãos do terceiro setor, escolas e instituições em geral, para prestarem serviços de avaliação psicológica. Nesse sentido, faz parte desse processo de avaliação a utilização de testes psicológicos como uma das ferramentas e/ou instrumentos que fornecem dados referentes à amostra de comportamento que, segundo Anastasi (2000), é, também, uma medida objetiva e padronizada do fenômeno psicológico. Para serem válidos e precisos, os testes psicológicos devem passar por estudos que comprovem suas qualidades psicométricas, bem como devem atender a determinadas especificações que garantam o reconhecimento e a credibilidade por parte da comunidade científica e de leigos (Anastasi, 1977; Adánez, 1999; Pasquali 2001).

No Brasil, a Resolução do CFP nº 002/2003 estabelece os requisitos mínimos essenciais aos testes disponibilizados para uso profissional (Conselho Federal de Psicologia, 2003). Essa normatização determina que, para ser considerado teste psicológico, o instrumento deve ser construído de acordo com os princípios referidos pela Psicometria e pelos documentos existentes em âmbito internacional da *International Test Commission* (ITC, 2000), *American Educational Research Association American Psychological Association*, *National Council on Measurement in Education* (AERA, APA & NCME, 1999) e da *Canadian Psychological Association* (CPA, 1996).

O *Standards for Educational and Psychological Tests* (1999) refere-se aos padrões para construção, validação e documentação dos testes utilizados em avaliação psicológica e

considera a validade a propriedade mais importante na construção de um teste. Esse dado é relevante e reforça a tese de que nenhum instrumento pode ser publicado sem um manual que reserve uma parte e/ou capítulo de estudos da validade.

Tal propriedade, a validade, refere-se ao grau em que o instrumento mede aquilo que se propõe medir, isto é, verifica em que medida ele atende aos objetivos propostos (Anastasi & Urbina, 2000; Sisto, Codenotti, Costa & Nascimento, 1979). Para Cronbach (1996, p. 143) “a validade é uma investigação da correção de uma interpretação”, e, sob essa perspectiva, só tem sentido dentro de um contexto, relativo ao tempo e ao espaço. Em outras palavras, a validade precisa ser realizada com seriedade na tomada de decisões, em consonância com a respectiva população.

Nessa perspectiva, a escolha por um tipo de validade ou outro deve ser realizada em função do objetivo do teste. Desta forma, a validade de um teste pode ser determinada somente para uma decisão ou situação e provavelmente não o seja para outras. Como já afirmamos anteriormente, os padrões de validade foram propostos pela *American Psychological Association* (APA) que, em 1954, previa a validade de conteúdo, a validade preditiva e concorrente e a validade de construto, todas estabelecidas conforme os objetivos do teste. Num momento posterior, em 1966, tanto a validade preditiva quanto a validade concorrente transformaram-se em uma outra, a de critério. Na literatura atual, a APA considera que toda a validade é referente ao construto.

Nesse sentido, considerando a necessidade de que todo instrumento apresente estudos que evidenciam sua validade, e que estas devam ser acumuladas para se ampliar a compreensão dos escores dos testes, este estudo se insere na busca do aprimoramento dessa

qualidade, procurando fornecer evidências que ampliem as possibilidades de interpretações dos escores dos testes, neste caso específico, do Teste Gestáltico Visomotor de Bender. Mais precisamente, pretende-se com este trabalho buscar evidências de validade convergente-discriminante do Sistema de Pontuação Gradual do Teste (B-SPG). Estas medidas são normalmente correlacionadas, verificando-se se, realmente, os instrumentos convergem em suas avaliações. Uma outra maneira de verificar essa validade seria separando grupos extremos na medida do critério, no intuito de verificar se são diferenciados em razão das pontuações do instrumento que se quer validar (Anastasi & Urbina, 2000).

Pretende-se também, reforçar a necessidade de buscar evidências de validade de um instrumento de uso constante pelos profissionais da área de clínica e educacional. São muitas as queixas apresentadas pelos professores sobre a aprendizagem do aluno, e dessa forma, verifica-se a necessidade de avaliar determinados aspectos que podem estar direta ou indiretamente relacionados à aquisição da referida aprendizagem. E assim, alguns instrumentos, em função da escassez de material na área de avaliação psicológica no país, acabam por ser utilizados maciçamente entre os psicólogos.

Pesquisas revelam que o modelo de correção mais utilizado no Brasil é o proposto por Koppitz (1989), embora estudiosos como Sisto, Noronha e Santos (2005) tenham levantado questionamentos a respeito da adequação dos critérios de pontuação por ela propostos. Há, todavia, esforços no sentido de se buscar outros sistemas de correção para captar os aspectos maturacionais no desenvolvimento infantil (Shapiro & Simpson, 1995; Lacks, 2000). Da mesma forma, Chan (2001) apontou sobre a dificuldade de comparação

dos resultados de Bender com crianças chinesas de 4,5 anos a 8,5 anos, quando pontuadas segundo a classificação norte-americana. Como resultado o autor demonstrou a necessidade de padrões apropriados às crianças avaliadas.

Nessa mesma direção, no que se refere ao sistema de pontuação do teste, é preciso ressaltar que, embora Bender, a autora do instrumento original, não tenha criado um procedimento sistemático de correção de seus cartões; diferentes autores em todo o mundo apresentaram tentativas de atribuir pontos a determinados indicativos de sinais neurológicos, de características afetivo-emocionais e de organização grafo-perceptiva. Esses componentes, de uma forma ou de outra, interferem no processo de aprendizagem infantil (Field, Bolton & Dana 1982). Com isso, a presente pesquisa buscou verificar se o Sistema de Pontuação Gradual do Bender, publicado recentemente no país (Sisto, Noronha & Santos 2005) apresenta correlação ao ser aplicado com outros testes; que contemplem construtos teoricamente relacionados, quais sejam: aprendizagem da leitura e da escrita. Pressupõe-se que a criança, para apresentar um bom desempenho nessas áreas, requer maturidade da percepção visomotora Koppitz (1989).

Em seus estudos, Bender (1974) argumenta que as crianças que apresentaram um bom desempenho no teste Bender também eram bem sucedidas na escola. Por isso, o teste tem sido apontado como um instrumento útil para o diagnóstico diferencial de crianças com problemas de aprendizagem, desde que o desenvolvimento percepto-motor esteja adequado para a faixa etária esperada; tem sido utilizado também como requisito para o desenvolvimento de habilidades acadêmicas, como relatam Sisto, Noronha e Santos (2005). Contudo, embora amplamente utilizado nacional e internacionalmente, é necessária a busca

de evidências de validade para muitos dos usos do Bender, assim como o desenvolvimento de novas propostas de sistema de correção, a fim de sanar as limitações na interpretação das pontuações obtidas.

Dessa maneira, os questionamentos realizados pelos pesquisadores podem contribuir para a área de psicometria, no que se refere à determinação de evidências de validade e de precisão e, principalmente, realçar a necessidade de deixar mais claros os limites dos testes usados e suas possibilidades de interpretação. Assim, pesquisadores e profissionais de psicologia contarão com instrumentos mais adequados e, em consequência, poderão produzir diagnósticos mais confiáveis (Sisto, Noronha & Santos, 2005).

É nesse sentido que se pretende realizar o presente trabalho, com a preocupação em atender aos requisitos mínimos e aceitáveis de validade e precisão, no intuito de garantir resultados confiáveis. O objetivo do pesquisador, ao construir ou validar um instrumento, é contribuir para o aprimoramento, bem como estimular futuras pesquisas com os procedimentos empregados na avaliação psicológica. O Brasil, retomando o que foi dito em relação à quantidade de testes, conta com um número reduzido de material sobre a avaliação da aprendizagem e da percepção visomotora. Desta maneira, as pesquisas referentes ao Teste de Bender, ainda encontram-se incipientes. Fazem-se necessários outros estudos, para que o tornem apropriado para o uso, e, em consonância com os parâmetros psicométricos, exigidos pelo Sistema de Avaliação de Testes Psicológicos (SATEPSI), criado pelo Conselho Federal de Psicologia (2001).

Feitas estas considerações iniciais acerca da presente pesquisa explicita-se, a seguir, a estrutura deste trabalho. O capítulo I contempla o histórico do Teste Bender Gestáltico,

registrando-se as questões relacionadas aos princípios que nortearam a sua construção e os estudos voltados às principais formas de correção do teste, bem como apontando as mais relevantes pesquisas realizadas para detectar problemas de aprendizagem em crianças. Contempla também algumas pesquisas relativas à utilização do teste Bender para a predição do desempenho escolar, para o diagnóstico de problemas de leitura e para a avaliação de problemas emocionais em crianças e adultos.

Ao lado disso, apresentam questões sobre a utilidade do teste Bender, enquanto instrumento de diagnóstico diferencial de pacientes psiquiátricos adultos, além de estudos que correlacionam o Teste de Bender com a inteligência. Trata, ainda, das pesquisas estrangeiras e nacionais com o uso do Teste Gestáltico Visomotor de Bender. Tecem considerações sobre a validade e a precisão do instrumento, sobretudo apontando algumas limitações na utilização do modelo proposto por Koppitz (1975) para a avaliação de crianças brasileiras. E mencionam estudos relacionados aos questionamentos sobre os critérios de avaliação.

No capítulo II serão abordados os parâmetros psicométricos, quais sejam: a precisão, validade e padronização dos testes psicológicos, e serão feitas considerações sobre pesquisas com o sistema de pontuação Bender-SPG. Os objetivos serão descritos após o capítulo II e o Método, que compõe o capítulo III, tratará da descrição dos participantes e dos instrumentos que foram utilizados na pesquisa, tais como o Teste de Bender-SPG, a Escala de Avaliação de Dificuldades na Aprendizagem da Escrita (ADAPE), o Teste de Cloze e o Teste de Reconhecimento de Palavras.

Em seguida, serão relatados, no capítulo IV, os resultados da pesquisa, acrescidos da discussão dos dados obtidos, fazendo-se referência às estatísticas descritivas dos instrumentos, à correlação de Pearson entre os instrumentos, idade e escolaridade, à análise de grupos extremos, à precisão do Bender-SPG, e as diferenças de médias quanto às variáveis. Por fim, encontram-se as Considerações Finais, que contempla o capítulo V, e tecem argumentações em função dos objetivos propostos, propõem-se novas pesquisas e apontam as dificuldades encontradas durante o processo de coleta e análises, correlacionado-as aos resultados obtidos. Na seqüência, apresentam-se as Referências no capítulo VI e os Anexos, capítulo VII.

CAPÍTULO I

TESTE BENDER GESTÁLTICO: HISTÓRICO E PRESSUPOSTOS TEÓRICOS

O Teste Visomotor de Bender, construído com base na Psicologia da *Gestalt* e seus estudos sobre a percepção, foi criado pela psiquiatra americana Laretta Bender, em 1938. A autora interessou-se pelos estudos da Psicologia da *Gestalt* a partir de 1932, quando trabalhava numa clínica para doentes mentais em Maryland, com o objetivo de verificar se a percepção de esquizofrênicos e deficientes mentais obedecia aos princípios postulados por Wertheimer (1923). Este, sendo um dos precursores da *Gestalt* e fundador da escola de Berlim, juntamente com Koffka e Köhler, definiu numerosas leis de organização perceptual e ilustrou-as em mais de 30 configurações. Ainda neste sentido, Wertheimer (1923) e seus colaboradores escreveram um artigo sobre os fenômenos perceptivos, entre eles, o movimento aparente e pós-imagens. Esses experimentos de gúestaltés visuais foram aplicados em crianças e adultos, débeis mentais e outros grupos de sujeitos.

A psicologia da *Gestalt* deu um novo impulso à ciência psicológica e contribui com estudiosos e clínicos da época, numa perspectiva diferente do aprendizado entre as relações do todo e de suas partes, demonstrando que a percepção não pode ser interpretada como uma junção de sensações singulares. Bender (1938) selecionou nove das configurações sugeridas pelos autores da *Gestalt* e, ao invés de solicitar aos pacientes que descrevessem o que viam na figura, como Wertheimer (1923), sugeria que os modelos apresentados fossem copiados.

A autora utilizou-se desta prática, com pessoas normais, crianças e adultos, e dessa experiência clínica surgiu o Teste de Bender, transformando uma tarefa que era apenas

visual em visomotora. Analisando as reproduções dos sujeitos, Bender (1938), concluiu que o produto final da cópia da figura é um padrão visomotor modificado, em relação ao padrão original, pela ação de mecanismos integradores do indivíduo que vivenciou a experiência. Assim, o padrão de respostas depende da configuração do estímulo e do estado de integração do indivíduo (ou do paciente) que, em condições patológicas, como nas doenças mentais, fica comprometido.

Para melhor compreensão do teste, Bender (1938) sugeriu que fossem retomados os princípios básicos da organização do campo visual propostos por Wertheimer (1923). Tais princípios foram formulados sob a forma de leis, como as da *proximidade*, da *semelhança*, da *uniformidade*, da *configuração objetiva*, e da *direção* ou *continuidade*. Na *lei da proximidade*, o indivíduo tende a perceber como unidades os elementos com os mesmos contornos, enquanto na da *semelhança* há uma tendência a perceber como agrupados elementos parecidos.

A crença de que a organização da disposição dos elementos é responsável pela determinação do todo é a base da *lei da uniformidade*. Ao lado disso, na *configuração objetiva*, um elemento é parte de uma seqüência, e a constelação resultante da seqüência é diferente daquela produzida por outra. Assim, um arranjo ambíguo se torna perfeitamente definido e inequívoco quando dado como parte de uma hierarquia. Por fim, na *lei da direção ou continuidade*, certos arranjos são mais fortes que outros devido à justaposição das partes.

Outras propriedades relevantes ao teste são: simetria, equilíbrio e fechamento. Com base nesses postulados, Wertheimer (1923), formulou a lei fundamental da *Gestalt*,

intitulada *Boa Forma*. Para o autor, quando ocorre conflito entre duas formas prevalece a mais regular, simétrica simples, enfim, a mais fortemente organizada, isto é, a que tem maior estabilidade no campo perceptivo, melhor pregnância.

A *Gestalt* encontra, nesses fenômenos da percepção, as condições para a compreensão do comportamento humano, de tal maneira que a forma como se percebe determinado estímulo é responsável pela resposta emitida pelo sujeito. Ainda nesse sentido, ou seja, o da função gestáltica visomotora, Bender (1969) observou que esta é sujeita ao processo de maturação ou crescimento, e que os primeiros desenhos das crianças se apresentam como puro jogo motor. Os rabiscos resultantes são pobres de sentido e se repetem diferentemente do modelo apresentado para copiar.

Baseada, portanto, em sua experiência clínica com o teste, Bender chegou a ampliar alguns conceitos da Teoria da Forma. Ao lado disso, Tolor e Schulberg (1963, p. 12), descreveram suas inovações:

A função gestáltica visomotora, por influência da maturação (diferenciação endógena), passa de um estado global, primitivo e indiferenciado para a construção de uma figura sobre um fundo. Este processo maturacional é função do tempo, da mesma forma que a ação. A ação resulta na experiência da percepção e é função do organismo como um todo, incluindo o fator motor.

Para os autores, a construção de um padrão percebido tem como base biológica os movimentos de rotação circular do campo óptico, de tal sorte que a imagem corporal é uma *gestalt* perceptivo-motora e, como tal, pode ser refletida por experiências perceptivo-

motoras como as evocadas pelo teste. Bender (1974) escolheu a Figura A (Anexo 1) para iniciar o teste, considerando-se a simplicidade de reprodução de ambas as partes (círculo e quadrado), fato que permite aos sujeitos identificar com facilidade as suas *gestaltes*. No caso da Figura 1, esta seria percebida como uma série de pares de pontos, tendo um ponto solto em cada extremo. A Figura 2 se percebe, segundo Wertheimer (1923), como uma série de linhas oblíquas compostas de três unidades, dispostas de maneira que as linhas estão inclinadas desde a esquerda acima até a direita abaixo.

Na Figura 3, embora o desenho representado se assemelhe a uma ponta de flecha, a *gestalt* das proximidades está presente. No caso da Figura 4, a estrutura está relacionada como uma das unidades estabelecidas pelo princípio da continuidade da organização geométrica, por meio da representação de um quadrado aberto na borda superior e de uma onda interceptando um dos seus vértices. O mesmo princípio rege as Figuras A e 5, enquanto a Figura 6 se encontra formada por linhas sinuoidais de diferentes longitudes de onda que se cruzam obliquamente. As Figuras 7 e 8 são duas configurações compostas pelas mesmas unidades, mas pouco se percebe, enquanto que na Figura 8 prevalece o princípio da continuidade das formas geométricas, na Figura 7, elas se tangenciam (Bender, 1974).

O Teste de Bender, como outros instrumentos de avaliação psicológica, passou por amplos estudos. A busca de sua história, de sua origem e dos principais princípios que nortearam sua construção, faz-se necessária, bem como o conhecimento dos principais sistemas de correção apontados pelas pesquisas, envolvendo amostras diferenciadas em relação ao gênero, idade, patologias, entre outras.

PRINCIPAIS SISTEMAS DE CORREÇÃO E ATRIBUIÇÕES DO TESTE BENDER

Bender (1969) adotou dois tipos de enfoque para a avaliação dos protocolos, um evolutivo e outro clínico. No primeiro caso, descreveu o processo de maturação da função gestáltica visomotora em crianças, apresentando um diagrama com ilustrações típicas do desempenho dos 3 aos 11anos. No segundo, efetuou uma avaliação clínica dos desenhos de pacientes adultos, de acordo com pressupostos teóricos e experiência prática. Em ambos os casos, a apreciação tende a ser subjetiva e baseada numa impressão global sobre a prova. Considerando a preocupação dos psicólogos com maior objetividade na análise dos protocolos, surgiram novos métodos de avaliação dos modelos, sendo alguns destinados a adultos e outros a crianças.

Entre os sistemas de avaliação para adultos, salienta-se o de Billingslea (1948), o primeiro método objetivo e quantitativo de avaliação do teste. No entanto, os achados da literatura apontam que o sistema de Pascal e Suttel (1951), o de Hutt (1960) e o de Koppitz (1989) são os mais conhecidos, sendo o da Koppitz o mais estudado e utilizado entre todos eles. Hutt e Briskin, em 1960, utilizaram o Bender como um teste projetivo não-verbal e interpretaram os desenhos obtidos à luz da teoria psicanalítica, restringindo a aplicação da prova a sujeitos com a função gestáltica visomotora plenamente desenvolvida. O sistema Hutt (1969) também tem sido considerado eficiente para o diagnóstico diferencial em psicopatologias. Lownsdale, Rogers e McCall (1989) e Lee e Oh (1998) entre outros, realizaram pesquisas em sujeitos com diagnósticos diferentes como depressão,

esquizofrenia e demência e encontraram como resultado nos testes de personalidade, índices de correlação significativos.

O sistema de Pascal e Suttell (1951) é o mais aceito e utilizado com pacientes psiquiátricos. Os autores criaram um sistema para a estimativa do grau de ajustamento emocional do indivíduo; estudaram indivíduos na faixa de 15 a 50 anos de idade, de inteligência normal com maturidade e capacidade para perceber e reproduzir, sem erro, as figuras do Bender. Feito isso, consideraram o desempenho do sujeito no teste como um reflexo de suas atitudes frente à realidade, sendo a capacidade de desenhar figuras uma função integradora, ou da sua “força do ego, e sugeriram que a magnitude do escore do Bender está relacionada com a gravidade da perturbação emocional. E, ainda, reconheceram o efeito da maturação no desempenho do teste e recomendaram o cuidado quando aplicado em crianças menores de 9 anos; para eles, o instrumento não deve ser aplicado em crianças menores de 6 anos. Difundido também em outros países, o sistema de Pascal e Suttell apresentou evidências de validade que o torna adequado no diagnóstico diferencial, como verificado na pesquisa de McCann e Plunket (1984).

Em sua pesquisa, os autores anteriormente citados aplicaram o Teste Gestáltico de Bender em 30 sujeitos com queixa de psicose Korsakoff, 30 com esquizofrenia paranóide e 30 sujeitos normais, para controle. A aplicação ocorreu de quatro formas diferentes. Na primeira, os cartões eram apresentados por 10 segundos e removidos, imediatamente após esse processo era solicitada uma reprodução da figura. Na segunda forma, utilizou-se o método padrão de aplicação do teste. Na terceira, solicitou-se aos sujeitos que desenhassem as figuras com a mão oposta à de sua preferência. E, por último, apresentaram-se os cartões

e os desenhos realizados no modelo padrão e foram perguntadas aos participantes as diferenças entre eles. Concluiu-se que todas as quatro formas de aplicação evidenciaram validade concorrente em relação às patologias estudadas. Ao lado disso, a quarta forma demonstrou potencial para o aperfeiçoamento da validade preditiva do Teste de Bender com a psicose de Korsakoff e esquizofrenia paranóide.

Outros estudos foram realizados com o intuito de buscar evidências de validade e precisão para alguns dos sistemas relacionados aos aspectos maturacionais do Bender. Nesse sentido, vale destacar o modelo de correção de Koppitz (1975) que foi construído por meio de uma escala de maturação para crianças e um sistema de pontuação. A população utilizada para a pesquisa pertencia a escolas públicas dos Estados Unidos. A autora, diferenciou os desvios que refletem imaturidade ou disfunção perceptiva daqueles que está relacionado a problemas emocionais.

Os resultados foram obtidos a partir da aplicação e correção dos protocolos, dentre os quais se encontravam algumas crianças excepcionais, outras com problemas emocionais, lesão cerebral, dificuldades de aprendizagem, deficiência mental e outras ainda que frequentassem escolas de nível primário. Este método, pela frequência em seu uso, gerou polêmica e muitos estudos.

Dentre eles, destacam-se o de Mazzeschi e Lis (1999), que estudaram a adaptação do sistema de correção proposto por Koppitz (1989) para amostras italianas e as diferenças de média por idade, bem como verificaram se existe relação entre diferenças culturais e desenvolvimento psicomotor nos vários grupos étnicos. Observaram-se estruturas familiares diferentes nas duas regiões: no norte, os pais compartilham a educação dos

filhos, valorizando a autonomia deles, enquanto que, no sul, a educação é legada à responsabilidade da mãe, ou seja, esta representa a figura central para o desenvolvimento dos filhos.

Os dados foram coletados em uma população de 1.065 sujeitos sendo 527 do sexo feminino e 538 do sexo masculino, com idades entre 3 a 11 anos, e localizada em diferentes regiões da Itália. Os resultados da análise da variância apontaram que não houve diferenças entre as crianças das regiões norte e sul da Itália em relação ao sexo, região, idade e cultura. Nesse sentido, as autoras sugerem cautela na interpretação desses resultados, considerando que os escores estão baseados em dados normativos derivados de outra população.

Nessa mesma direção, Weinstein (1983) realizou uma pesquisa sobre a diferença entre sexos utilizando-se do Teste de Bender, sistema de correção da Koppitz. Os sujeitos eram 60 crianças de ambos os sexos de uma escola pública americana. Os resultados revelaram que somente a categoria distorção de forma encontrou diferenças significativas no desempenho entre meninos e meninas, embora os meninos tivessem apresentado uma maior dificuldade na realização de alguns desenhos.

Numa outra pesquisa, o Teste de Bender foi utilizado na predição de desempenho em leitura e matemática. Os outros dois instrumentos utilizados foram o Wide Range Achievement Test (WRAT), para avaliar a leitura e a matemática, e a *Scale the Intelligence Wechsler for Children – Revised* (WISC-R). Os sujeitos foram 153 crianças com idades entre 6 e 12 anos. Foram comparados dois grupos de crianças, atrasadas e normais em relação à idade e à série escolar. Os participantes do primeiro grupo, concentraram os piores escores no Teste de Bender e na aquisição da leitura, além de apresentarem QIs mais

baixos. Dessa forma, o Bender pareceu ser mais útil na predição de desempenho em leitura e matemática de crianças atrasadas (Nielsen & Sapp, 1991).

Mais recentemente, outros sistemas surgiram, dentre os quais se destaca a versão modificada de Brannigan e Brunner (1989) para ser utilizada com crianças de 4 a 8 anos. Essa versão consiste no uso de 6 figuras das 9 utilizadas por Bender, que seriam as figuras A, 1, 2, 4, 6 e 8, com a função de avaliar a representação global de cada desenho. Em relação à pontuação, esta varia de 0 a 5 pontos com o escore máximo de 30 pontos. Quanto à validade e precisão, foram apresentados estudos posteriores que evidenciaram resultados estatisticamente significativos sobre o sistema no que se refere a esses quesitos. Dentre eles, apresentam-se os trabalhos de Brannigan e Brunner (1991), Brannigan e Brunner (1993) e de Chan (2001).

Entre os sistemas mais conhecidos no Brasil, alguns foram estudados em pesquisas, como é o caso do sistema de Santucci e Galifret-Granjon (1968) o de Santucci e Pêcheux (1981) o de Koppitz (1989), e o de Clawson (1999). Embora tenham sido citados vários sistemas de avaliação para a verificação das dificuldades específicas em crianças, um dos mais utilizados em nosso meio é o *Developmental Bender Test Scoring System*, criado por Koppitz em 1975, já citado anteriormente e com o objetivo de avaliar a maturidade neurológica da criança para o diagnóstico de lesão cerebral.

Vale retomar que o sistema Koppitz (1975) foi elaborado a partir de uma escala de maturação para o Bender e um sistema de pontuação para crianças pertencentes a 46 classes de 12 escolas públicas, dos Estados Unidos, diferenciando os desvios que refletem imaturidade ou disfunção perceptiva daqueles relacionados a problemas emocionais. Os

resultados foram obtidos a partir da aplicação a 1.104 crianças, de 5 a 10 anos e 11 meses. A escala compreendia de 20 itens inicialmente: distorção da forma, rotação, rasuras, omissões, ordem confusa, superposição de figuras, compressão, segunda tentativa, perseveração, círculos ou traços em vez de pontos, linha ondulada, desvio na inclinação, traços ou pontos em vez de círculos, achatamento, número incorreto de pontos, quadrado e curva que não se juntam, ângulos na curva, omissão ou adição de ângulos e moldura nas figuras distribuídas em quatro categorias de erros, a saber. A Distorção da Forma (todas as nove figuras) está relacionada à desproporção entre as partes, números incorretos de pontos, perda da forma do desenho, linhas em lugar de pontos, falta de integração das partes. Em acréscimo, por Distorção de Forma, entende-se o desrespeito às formas estruturais do desenho, de tal modo que pontos, linhas, retas, curvas e ângulos são copiados, sem precisão.

A Rotação diz respeito à modificação da orientação do desenho em relação ao estímulo, ou seja, a figura somente será pontuada quando a alteração for de 45 graus ou mais no eixo da figura. Perseveração (Figuras 1, 2 e 6), com mais de 14 pontos ou colunas e mais de cinco curvas sinusoidais, a categoria A Perseveração diz respeito ao aumento do número de elementos reproduzidos em comparação com a figura apresentada. A Integração (Figuras A, 2, 3, 4, 5, 6, e 7), relacionada à falha na união das subpartes. A categoria Integração está relacionada à perda da configuração em função do fracasso na união das partes, omissão, acréscimo ou substituição de elementos, perda da posição relativa ou da modificação das partes estruturais da figura.

No total, o teste inclui 30 itens, e as alterações em cada uma destas categorias são pontuadas. Uma alta pontuação indica um baixo desempenho, no entanto, uma baixa pontuação está relacionada a um bom desempenho. Percebeu-se também que conforme a idade aumenta, o número de distorções diminui o que sugere o amadurecimento da habilidade visomotora. Em razão da dificuldade proporcionada por este tipo de avaliação, a autora aponta para que sejam computados apenas os casos em que aparece desvio evidente e alerta para que, em caso de dúvida, não se pontue.

Os resultados de Koppitz (1975) apontaram a maturidade da criança em termos de percepção visomotora, além de fornecer indícios para o diagnóstico de lesão cerebral. Após a aplicação individual, correção e análise de todos os protocolos, a autora, concluiu, que a medida é sensível para captar o desempenho percepto-motor dos 6 aos 10 anos, idade por idade, o que demonstra evidências de validade em relação ao aspecto desenvolvimental.

A precisão da escala maturacional foi investigada por meio da técnica de teste-reteste. Foram encontradas correlações entre 0,55 e 0,66 no nível de $p=0,001$, resultados considerados estatisticamente significativos. No entanto, pautando-se nos critérios de Prieto e Muñiz (2001), essas correlações podem ser consideradas apenas adequadas, mas com algumas carências. Koppitz (1989) verificou que a maioria dos meninos e meninas de 9 anos é capaz de reproduzir as figuras sem erros grosseiros e que, após essa idade, a ausência de erros indica que a habilidade visomotora está em conformidade com o esperado para a sua faixa etária. A autora propõe que a medida seja utilizada para identificar habilidade visomotora imatura ou deficiente.

Outra análise realizada por Koppitz (1989), foi o significado do fator tempo para a realização das figuras no Teste de Bender, bem como o tempo médio utilizado para a sua aplicação. Com base na análise do desempenho de 339 crianças residentes nos Estados Unidos, a autora verificou que uma criança com idade entre 5 e 9 anos precisa de um tempo médio de 6 minutos e 30 segundos para a reprodução das 9 figuras. O resultado apontou que a reprodução muito rápida das figuras está associada ao baixo desempenho visomotor, à falta de concentração ou esforço na realização da tarefa ou rendimento escolar inferior. A autora complementa que protocolos desse tipo são mais comuns entre as crianças com lesões cerebrais. Ao lado dessa escala, Koppitz (1989) propôs uma análise do ajustamento emocional da criança, por meio de uma série de dez indicadores, para diferenciar as crianças normais das emocionalmente perturbadas. Tal análise não será aprofundada neste estudo considerando que os aspectos nele abordados estão relacionados aos de natureza cognitiva.

Outro sistema verificado na literatura, foi o de Santucci e Pêcheux (1981) que sistematizaram os critérios de correção para a Prova de Organização Perceptiva para crianças com idade entre 6 e 14 anos. Esses critérios foram elaborados no sentido de pesquisar uma possível deficiência na organização perceptiva de crianças com atraso escolar, bem como para verificar a relação entre a deficiência mental e a organização percepto-motora. Nesse sentido, os autores selecionaram cinco das nove figuras propostas por Bender, que são as figuras A, 2, 3, 4 e 7 e realizaram um estudo relacionado à evolução genética das reproduções. Quanto à classificação, foram considerados os elementos forma,

número, relação de continuidade e separação, orientação geral, precisão dos alinhamentos e exatidão das dimensões.

Quanto à avaliação dos desenhos, foram atribuídos pontos de acordo com três níveis de sucesso na técnica de Santucci-Pêcheux: 3 pontos para os aspectos raramente reproduzidos pela maioria das crianças de 6 anos, 1 ponto para os aspectos bem reproduzidos por muitas crianças desde os 6 anos e 2 pontos para os casos intermediários (Cunha, 2000). Esses estudos foram realizados em Paris, com uma amostra de aproximadamente 300 crianças, não sendo encontrados estudos no Brasil que abordassem esse sistema de correção.

Em relação aos sistemas utilizados com crianças, devem ser citados o de Santucci e Galifret-Granjon (1968), que utilizaram apenas 5 figuras dentre as 9 propostas por Bender (1938) e cujos resultados indicam os ângulos, a orientação e a posição relativa da cópia de figuras como os três fatores que refletem ou não a reprodução exata das mesmas. Com o objetivo de estudar a influência de três variáveis, idade, sexo e nível de escolaridade, Machado (1978) pesquisou o sistema de Santucci e Galifret-Granjon numa população de 529 crianças brasileiras de ambos os sexos, com idades entre 7 e 10 anos. Utilizando-se dos cinco modelos propostos pelos referidos autores, observou que o sistema era sensível para captar diferenças de média em relação à idade. Quanto ao nível socioeconômico (NSE), os resultados revelaram que as crianças com NSE inferior apresentaram um acentuado atraso no desenvolvimento visomotor. Não houve diferença significativa em relação ao sexo.

Ainda em relação aos sistemas de correção do Bender, Posada (2002) também buscou desenvolver uma forma alternativa de correção, elaborando um sistema de classificação das respostas fundamentado nos princípios piagetianos de construção e representação espacial. As análises estatísticas preliminares de sua classificação indicaram parâmetros psicométricos significativos e vantagens em relação aos métodos clássicos de correção de Koppitz (1989) e Santucci e Pêcheux (1981).

Uma outra forma de correção também muito utilizada foi a proposta por Clawson (1999), que estabeleceu indicadores de perturbação emocional. Tal sistema não será abordado na presente pesquisa, por não avaliar os aspectos cognitivos da organização percepto-motora. Uma revisão da literatura realizada na década de 60 revelou a publicação de mais de 130 obras sobre o Teste Bender (*Field, Bolton & Dana 1982*). Muitas delas referiam-se à modificação na forma de aplicação, outras se relacionavam com o diagnóstico diferencial de pacientes psiquiátricos adultos, e outras, ainda, correlacionavam o Bender com a inteligência. Aproximadamente um quinto de todas as publicações sobre o teste foi dedicado às crianças, principalmente depois de 1955, evidenciando uma preocupação crescente sobre o valor do instrumento para a faixa etária infantil.

Ainda no que se refere ao uso do instrumento, o Teste de Bender tem sido muito utilizado em crianças. Dessa forma, o mesmo é indicado para detectar problemas de aprendizagem, para a predição de desempenho escolar, para o diagnóstico de problemas de leitura e de escrita, para a avaliação de problemas emocionais e das necessidades de psicoterapia, suspeita de lesão cerebral e a sua relação com o retardo mental (*Salvia & Ysseldyke, 1991*).

Nessa perspectiva, pode-se citar o trabalho de Bandeira e Hutz (1994), que investigaram o grau de predição do rendimento escolar na primeira série, por meio da aplicação dos testes de Bender, Raven e o Desenho da Figura Humana. Os resultados mostraram correlações significativas entre os três instrumentos e o rendimento escolar, apesar de somente o Bender e os itens evolutivos do Desenho da Figura Humana terem contribuído para explicar a variância do rendimento escolar.

Nessa mesma direção, Kroeff (1988, 1992), utilizando-se do modelo de correção proposto por Koppitz (1989), apresentou normas específicas para escolas públicas e privadas brasileiras e analisou a diferença de resultados de acordo com a série, cujos resultados apontaram para a necessidade de revisão do modelo proposto pela autora. Na pesquisa de Moura (1982), realizada com 160 sujeitos de 3 a 6 anos, avaliaram-se as diferentes *performances* dessa faixa etária, ficando evidente o caráter evolutivo na reprodução de figuras geométricas, embora não tenham sido essas as idades inicialmente estudadas por Koppitz (1975).

O sistema Koppitz de pontuação foi usado por Sisto, Bueno e Rueda (2003) que estudaram as relações entre traços de personalidade e percepção visomotora. Os sujeitos foram 344 crianças, entre 6 e 12 anos, de ambos os sexos. Os resultados apontaram que houve uma integração menor associada a maiores pontuações em neuroticismo, enquanto que maiores pontuações em distorção foram associadas a maiores pontuações em psicoticismo.

Os estudos citados assinalam a diversidade encontrada na aplicação do Teste de Bender a crianças e, no próximo tópico, serão abordados os trabalhos relacionados ao

instrumento, em âmbito estrangeiro e nacional. Apesar das várias publicações, faltam ainda, no Brasil, pesquisas de evidências de validade do Bender. A maior parte das publicações está relacionada a experiências de aplicações e correções, segundo os modelos de Koppitz (1989), Clawson (1999) e Santucci (1981), para fins de diagnóstico clínico.

ESTUDOS RELACIONADOS AOS QUESTIONAMENTOS SOBRE OS CRITÉRIOS DE AVALIAÇÃO

Mesmo sendo largamente utilizado e havendo suscitado muitas pesquisas que demonstravam consonância com os resultados apontados por Koppitz (1975), o sistema também foi alvo de muitas críticas. Assim, vários estudos que identificaram problemas relacionados ao sistema de correção e aos pressupostos teóricos serão relatados, apresentando-se inicialmente os realizados no estrangeiro e depois os brasileiros.

Buckley (1978) revelou que faltam evidências de validade para muitos dos usos indicados pelo sistema, como, por exemplo, para a predição de rendimento escolar ou problemas neurológicos, que se mostra frágil. Afirmou, ainda, que as normas de Koppitz não eram satisfatórias para a avaliação de desempenho de crianças vindas de minorias ou com algum tipo de privação. Outros pesquisadores como Taylor, Kauffman e Partenio (1984) Shapiro e Simpson (1995) levantaram dúvidas sobre o valor dos critérios de correção utilizados, sugerindo que há limitações na interpretação da pontuação obtida, as quais, conseqüentemente, comprometem a sua utilização para diagnosticar problemas neuropsicológicos e os aspectos relacionados ao desenvolvimento infantil. Rie, Rie, Stewart e Rettemnier (1978) identificaram a existência de dois fatores neurocognitivos distintos, relacionados à integração verbal-motora e visomotora, em crianças com problemas de aprendizagem.

Ainda no que concerne aos estudos que criticam o uso do Bender para a avaliação do desempenho escolar ou das disfunções neurológicas, encontra-se o de Blanchard (1998).

Este teve como objetivo identificar em 387 pacientes de grupos clínicos distintos (adolescentes institucionalizados, dementes senis, pacientes psiquiátricos, entre outros) qual figura ou combinação de figuras seria mais útil para um diagnóstico rápido de disfunções neurológicas. Foram utilizados diferentes métodos de correção, e os resultados apontaram a Figura 7 como a que apresentou correlações mais altas e consistentes em todos os métodos e para todos os grupos de pacientes, o que revela que esta seria a figura mais indicada para um diagnóstico rápido e preciso de tais funções.

Nesse mesmo intuito, McKay e Neale (1985) realizaram um estudo envolvendo 195 crianças e confirmaram algumas das dificuldades do uso do sistema de correção e classificação de Koppitz (1975). Os resultados demonstraram que as categorias de distorção, círculos por pontos e linhas por pontos foram melhores que o escore global na predição de desempenho acadêmico, principalmente na leitura e na escrita.

Outros estudiosos também realizaram trabalhos com o intuito de verificar a validade do Bender como preditivo do rendimento acadêmico. Dentre eles, destacam-se Brannigan, Aabye, Baker e Ryan (1995), que compararam o sistema de classificação de Koppitz (1975), utilizando-se dos nove desenhos originais, com o sistema qualitativo de Brannigan e Brunner (1989), que usam apenas seis figuras. Os sujeitos foram 409 crianças dos primeiros quatro anos da escola básica, sendo os resultados obtidos em ambos os sistemas correlacionados com os obtidos no teste Metropolitano de Prontidão. Após análises, verificou-se que o sistema qualitativo apresentou-se mais sensível para a identificação de crianças com problemas de aprendizagem.

Da mesma forma, Vance, Fuller e Lester (1986) questionaram o valor preditivo do Bender em relação ao desempenho intelectual e acadêmico. Utilizando-se do sistema de correção de Koppitz e do *Minnesota Perceptual Diagnostic Test Revised* (MPDT-R), com crianças da área rural norte-americana, obtiveram resultados muito mais satisfatórios em relação à captação dos aspectos desenvolvimentais das crianças. Outro dado relevante foi o resultado do MPDT-R, que mostrou ser o melhor preditor do desempenho intelectual avaliado pelo WISC e do desempenho escolar em provas de matemática e linguagem.

Lesiak (1984) realizou uma revisão de 32 artigos publicados nos Estados Unidos a respeito de Teste de Bender entre 1962 e 1981 e os reuniu em três áreas em relação ao Bender: a utilidade do instrumento como preditor de leitura, a habilidade do teste em diferenciar bons e maus leitores e a relação entre o Bender e a realização em leitura. Os resultados apontaram que, como preditor de leitura, o Bender é adequado somente para indivíduos cursando a 1ª série, sugerindo-se, a partir dessa respectiva faixa etária, outros instrumentos para predição. A respeito do uso de diferentes amostras e dos procedimentos envolvidos nos estudos revisados, o desempenho dos indivíduos no Teste de Bender não guardou relações com a habilidade em leitura. Mediante as controvérsias observadas nos estudos, Lesiak (1984) sugeriu cautela na utilização da medida para a diferenciação entre bons e maus leitores, conforme alto ou baixo desempenho no Teste Gestáltico de Bender.

As pesquisas de Brannigan e Brunner (1991, 2002) e Moose e Brannigan (1997) consideraram que o sistema Koppitz (1975) apresenta falhas para captar diferenças entre as crianças da faixa etária para a qual o instrumento se destina (6 a 10 anos). Por conseguinte,

muitos autores têm trabalhado em versões modificadas do teste, ou mesmo em outros sistemas de avaliação mais qualitativa dos desenhos produzidos pelos examinandos.

Em razão dos problemas detectados, há autores que questionaram os critérios de correção utilizados por Koppitz (1975), principalmente com relação aos valores atribuídos às figuras, afirmando haver limitações na interpretação da pontuação obtida que, conseqüentemente, comprometem a sua utilização no diagnóstico neuropsicológico. Esse dado pode ser observado nos trabalhos de Weinstein (1983), que demonstrou que o sistema Koppitz não foi sensível o suficiente para detectar diferenças relacionadas ao sexo, principalmente em crianças menores. Participaram desse estudo 30 meninos e 30 meninas de 5 anos e três meses a 6 anos e dois meses. Os resultados apresentados não revelaram diferenças significativas, enquanto que, utilizando-se de outros sistemas, que consideraram aspectos diferentes dos de Koppitz, conseguiu-se detectar diferenças no desempenho dos meninos, significativamente inferiores ao das meninas da faixa etária pesquisada.

Chan (2001), em recente pesquisa, relatou a dificuldade encontrada de comparação dos resultados do Bender com crianças chinesas de 4 anos e 6 meses a 8 anos e 6 meses, quando pontuadas pelo sistema Koppitz (1975). O autor enfatizou, na conclusão dos resultados, a necessidade de se criar padrões apropriados às crianças dessa faixa etária. Em acréscimo, com base em amplo estudo realizado com 652 crianças norte-americanas, Taylor, Kauffman e Partenio (1984) questionaram a validade do sistema na identificação de características desenvolvimentais em crianças acima de 8 anos.

O sistema Koppitz também mostrou ter menor valor preditivo para as habilidades cognitivas quando aplicado em 103 crianças de idade pré-escolar. Esse resultado foi

verificado em um estudo longitudinal realizado por Aylward e Schimidt (1986), fruto da comparação entre os escores do Bender e os do VMI (Teste do Desenvolvimento da Integração Visomotora).

Em síntese, vários estudos foram realizados e apontaram, em seus resultados, as fragilidades do sistema Koppitz (1975). No Brasil, na mesma direção de estrangeiro, também foi possível encontrar resultados contraditórios em relação ao referido sistema. A fim de ilustrar estudos serão apresentados a seguir.

Pinelli Jr. e Pasquali (1990) procederam à avaliação dos parâmetros psicométricos do sistema de correção da Koppitz (1975) numa população brasileira, com base no método de análise fatorial para validação de construto. O Bender foi aplicado em 620 crianças entre 6 e 11 anos, divididas em dois grupos, de nível sócio econômico (NSE) diferente. Um alto, (Plano Piloto de Brasília), N=351 e outro baixo (Pedregal), N= 259. Os autores constataram uma estrutura multifatorial do Bender, ou seja, a existência de quatro fatores com consistência interna baixa, contrapondo-se a análise unifatorial do sistema de apuração de Koppitz (1979) que segundo eles, justificam novas pesquisas. Por outro lado, a análise da escala de Koppitz evidenciou que também aqui, existem problemas, uma vez que alguns itens podem ser eliminados sem prejuízo para a avaliação. Diante dos resultados, sugeriram que os itens precisariam ser organizados de outra forma, principalmente no que se refere ao aumento do número destes, em alguns fatores. Ainda, segundo os autores, são poucos os estudos que envolvem a análise fatorial do Bender no exterior, e inexistem com amostras brasileiras, pelo menos, até meados de 90, data de realização da pesquisa.

La Puente e Maciel (1984) por sua vez, defenderam uma avaliação qualitativa associada à quantitativa, uma vez que a primeira possibilitaria um diagnóstico mais completo. Eles pesquisadores discordaram do sistema de avaliação quantitativo do instrumento, afirmando que os critérios foram capazes apenas de uma avaliação em termos de normalidade/anormalidade, mas não da análise da dificuldade.

Já em estudo realizado por Brito e Santos (1996), encontraram-se resultados no sentido de que o Teste de Bender é válido para identificar aspectos psicopatológicos. No entanto, recomendaram cautela na utilização das normas propostas por Koppitz no que se refere à discriminação das variáveis relacionadas à idade e gênero aqui no Brasil. Nessa mesma perspectiva, Moura (1982) constatou que a percepção visual e a expressão motora, por serem fatores independentes, deveriam ser avaliadas separadamente, o que ficaria impossibilitado com a utilização do referido sistema.

Em um amplo estudo, Sisto, Noronha e Santos (2004) desenvolveram uma pesquisa com crianças das séries iniciais do ensino fundamental e buscaram analisar o critério de distorção de formas nas figuras do Teste de Bender, de acordo com o proposto por Koppitz. O estudo envolveu 314 crianças de ambos os sexos, estudantes das séries iniciais do ensino fundamental. Os autores perceberam a existência de correlação negativa entre idade e deformação em algumas figuras, o que já era esperado, considerando que, quando a idade avança, diminui a intensidade da deformação. No entanto, os resultados não foram os esperados para uma medida sensível ao caráter maturacional do desenvolvimento infantil, uma vez que os valores de alfa indicaram baixa consistência.

Num outro estudo, Sisto, Santos e Noronha (2004) analisaram os critérios de Koppitz para avaliar a integração das figuras do Bender e seu caráter evolutivo. Os participantes foram alunos de ambos os sexos, de 7 a 10 anos. Os resultados indicaram baixa correlação entre a média das idades e o número de erros, mesmo considerando que as correlações entre a análise da variância tenham se revelado estatisticamente significativas, apenas dois grupos etários foram diferenciados pelo teste de *Tuckey*.

Considerando a importância do Teste Gestáltico de Bender Sisto, Noronha e Santos (2005) entenderam como relevante a realização de estudos que o abordassem. Nesse sentido, disponibilizaram aos psicólogos brasileiros um novo sistema de aplicação e de correção adaptado às necessidades já apontadas por vários estudos nacionais sobre o instrumento. Assim sendo, os estudos psicométricos foram estabelecidos por meio de um conjunto de pesquisas realizadas e necessárias para a aprovação e comercialização de um instrumento, como é o caso do Bender-SPG (Sistema de Pontuação Gradual).

Nesse sentido, a próxima seção se destina a discutir os parâmetros psicométricos presentes na construção de testes psicológicos: os critérios de validade, precisão e padronização.

CAPITULO II

PARÂMETROS PSICOMÉTRICOS: PRECISÃO, VALIDADE E PADRONIZAÇÃO DOS TESTES PSICOLÓGICOS.

A primeira condição para que um teste sirva de base para levar a cabo inferências estabelece que a amostra de comportamento coletada seja precisa e que os erros cometidos na avaliação sejam ao menos aceitáveis, uma vez que nenhuma medição científica está totalmente isenta de erros (Muñiz, 2004). A fidedignidade ou precisão de um teste está relacionada à característica que ele deve possuir, ou seja, a de medir sem erros. Isso significa que o mesmo teste ou testes equivalentes, avaliando os mesmos sujeitos em ocasiões diferentes ou na mesma ocasião, devem produzir resultados idênticos, ou bastante semelhantes, de tal forma que a correlação entre as duas medidas se aproxime de 1,0 (Pasquali, 2003). Assim, a análise da precisão de um instrumento psicológico indica o quanto ele se afasta do ideal da correlação, uma vez que, quanto mais próximo de 1,0, menor é a possibilidade de erro.

Para Anastasi (2000), a precisão tem como base três aspectos principais, a saber: equivalência, estabilidade e homogeneidade. A equivalência se refere ao grau de acordo entre duas ou mais medidas feitas quase ao mesmo tempo. Para a estabilidade ocorrer deve ser feita uma distinção entre a repetibilidade da medida e a medida dos fenômenos que estão sendo avaliados. Por último, a homogeneidade trata de avaliar quão bem os diferentes itens de uma medida parecem refletir o atributo que pretendem aferir.

Sob essa perspectiva, a forma mais comum, de acordo com Anastasi (2000), de se verificar a precisão é por meio do teste-reteste, que se refere à reprodutibilidade dos valores de uma medida quando se avaliam os mesmos sujeitos duas vezes com o mesmo instrumento, com um intervalo de tempo entre as aplicações. Essa forma de análise apresenta alguns problemas, pois as medidas repetidas podem gerar um efeito de aprendizagem, pelo fato de os sujeitos enfrentarem duas vezes a mesma situação.

Outra possibilidade de verificação da precisão é por meio de formas paralelas, que se assemelha ao procedimento anterior, com o diferencial de que os instrumentos não são iguais, mas semelhantes, e que, sob essa perspectiva, a aplicação das duas provas pode ocorrer no mesmo dia. A correlação entre os dois instrumentos paralelos será a estimativa da precisão, pois indicará quão bem os itens rendem resultados similares.

Na precisão por consistência interna, por sua vez, a correlação é calculada aplicando-se o teste uma única vez. Nesse caso, são três as principais possibilidades de análise: duas metades (*split-half*), pela equação de *Kuder-Richardson*; ou pelo *Alfa de Cronbach*, uma extensão da técnica de *Kuder-Richardson*. Além disso, há a precisão por avaliadores que pode ser encontrada tendo-se uma amostra dos protocolos dos testes pontuados independentemente por dois ou mais examinadores. Os dois escores obtidos pelos testandos são correlacionados da maneira habitual, e o coeficiente de correlação resultante é uma medida da fidedignidade do avaliador (Anastasi & Urbina, 2000).

Conforme Ladesma, Adánez e Mora (2002), a técnica de análise de consistência interna, por meio do Alfa de Cronbach, é utilizada pelos pesquisadores para avaliação de medidas escalares, uma vez que, dentre as principais vantagens apresentadas, requer apenas

uma aplicação do instrumento de avaliação. Para Kline (1993), o valor mínimo aceitável para o coeficiente de Cronbach, é o de 0,80. Porém, Freire e Almeida (2001) analisaram o posicionamento de outros autores e sugeriram considerar intervalos de valores, destacando a necessidade da revisão do número de itens para torná-los heterogêneos quando o coeficiente for muito acima de 0,90; definiram como muito bom, entre 0,80 e 0,90; respeitável, entre 0,70 e 0,80; aceitável entre 0,65 e 0,70; indesejável, entre 0,60 e 0,65 e inaceitável, o coeficiente abaixo de 0,60.

De acordo com o definido pela Resolução 002/2003 do Conselho Federal de Psicologia (CFP, 2003), o valor mínimo aceitável para o coeficiente de precisão é de 0,60. Ainda no que se refere aos parâmetros psicométricos, a precisão e a validade são consideradas os aspectos mais importantes da medida. A comunidade científica tem concordado que a precisão é uma condição necessária, mas que ela não é suficiente para o estabelecimento da validade (Anastasi, 1967).

A mais recente edição dos *Standards for Educational and Psychological Testing*, publicado pela AERA, APA e NCME (1999), refere-se à validade como sendo a propriedade mais importante na avaliação de um teste. Ou seja, nenhum teste deve ser publicado sem que estejam presentes em seu manual, informações sobre os estudos de validade. Para Anastasi e Urbina (2000), a validade refere-se ao grau em que o instrumento mede aquilo que se propõe a medir e o quão bem faz isso.

O problema que se coloca é como comprovar que as inferências feitas a partir de um teste estão corretas, ou, mais objetivamente, como proceder para chegar a um processo satisfatório de validação das inferências feitas a partir de um teste. O resultado final de um

processo de validação não visa o estabelecimento simplista de que tal teste é válido, ou não.

O cerne da validação está em identificar quais caminhos devem ser traçados para se chegar a evidências empíricas ou teóricas, necessárias para se afirmar que são válidas determinadas inferências realizadas. Nessa direção os *Standards for Educational and Psychological Testing* (APA, 1980, p. 12) apontam: “Tradicionalmente as distintas formas de acumular evidências sobre a validade têm sido agrupá-las em validade de conteúdo, critério e construto” (). Os procedimentos de identificação do construto de um teste se referem a “extensão em que podemos dizer que o teste mede um construto teórico ou um traço” (Anastasi & Urbina , 2000, p.117). A validade de construto faz inferências sobre o grau em que uma pessoa possui uma determinada característica que, por suposição, possa aparecer no instrumento, por meio da definição de conceito.

Essas evidências podem ser obtidas de diferentes maneiras, como mudanças de desenvolvimento, que se baseiam na possibilidade dos escores aumentarem, de acordo com o aumento da idade; estrutura interna, na qual se busca verificar as relações entre os itens e a estrutura teórica hipotetizada quando da construção; e a evidência convergente-discriminante, que, por sua vez, refere-se à relação entre os resultados do teste e as outras medidas que avaliam um construto diferente ou similar. A esse respeito, a correlação de um teste de raciocínio quantitativo com notas subseqüentes em um curso de matemática seria um exemplo de validade convergente. Para o mesmo teste, a validade discriminante seria evidenciada por uma correlação baixa e insignificante com escores em um teste de compreensão em leitura, uma vez que a habilidade de leitura é uma variável irrelevante em um teste planejado para medir o raciocínio quantitativo.

Já no que compreende a validade de construto ou de conceito, ela é considerada por Pasquali (2001) como a etapa principal de validação de um instrumento psicométrico, uma vez que possibilita a verificação empírica da legitimidade da representação comportamental dos traços latentes. Por meio deste procedimento, é possível verificar se as pessoas com determinadas características comportam-se conforme a teoria supõe que elas comportar-se-iam e quais itens ou fatores são mais pertinentes ao construto (Cronbach, 1996).

De acordo com Pasquali (2001), a principal forma por meio da qual este tipo de validação é atestada seria utilizando-se da análise da representação comportamental do construto e da análise por hipótese. A análise da representação comportamental baseia-se em duas técnicas distintas, a saber, a análise de precisão do instrumento e a análise fatorial. A primeira consiste essencialmente em verificar a homogeneidade dos itens que compõem o teste. Assim, o escore total do teste torna-se o critério de decisão, e a correlação entre cada item e este escore total decide a qualidade do item. Se for alta a correlação, o item é retido. A segunda diz respeito à verificação da quantidade de construtos necessários para explicar as covariâncias dos itens. As correlações dos itens são explicadas, pela análise fatorial, como resultantes de variadas fontes, ou seja, os construtos ou traços latentes, que seriam as causas das covariâncias.

Sobre o tema, Muñiz (2004) destaca que, ao validar as inferências ou hipóteses dos testes, se valida também o modelo de hipóteses científicas. Todo processo de validação deve começar pela comprovação da pertinência dos conteúdos, o que envolve dois aspectos, a definição do construto e a sua correta representação no teste. Especialmente a definição do construto deve ser operacional e, para tanto, pode-se proceder de forma analítica e

racional, mediante a utilização de *experts* na temática, empregar técnicas estatísticas após a aplicação da prova ou, ainda, associar ambas as técnicas, obtendo-se diversos indicadores.

Ainda segundo o autor, as estratégias para reunir dados sobre os processos subjacentes às respostas das pessoas aos itens dos testes são muito variadas, apesar de se basearem nas respostas individuais. Essas estratégias consistem em, desde perguntar aos próprios sujeitos como eles procederam, observando os passos sucessivos que os conduziram ao resultado final, em utilizar observadores *expertos* e analisar, de forma experimental, os processos básicos e os componentes implicados na resposta de cada item.

Se um teste mede determinado construto e este é legítimo, ao aplicá-lo com outro teste de construto semelhante deve-se encontrar uma alta correlação, ou seja, os dois convergem. Ao contrário, dois instrumentos de construtos distintos, quando correlacionados, devem divergir, ou seja, é necessário diferenciar um construto do outro. Para evidenciar a validade de construto é imprescindível que haja o acúmulo gradual de informações de várias fontes. Qualquer dado que lance luz sobre a natureza do traço sob consideração, bem como sobre as condições que afetam seu desenvolvimento e suas manifestações, representa uma evidência apropriada para a sua validação (Muñiz, 2004).

Ao lado disso, Muñiz (2004) sugere que a evidência de validade de critério corresponde à relação entre os resultados do teste e outros critérios, indica a efetividade de um teste em predizer o desempenho de um indivíduo em atividades e requer uma medida de algo relacionado ao construto, mas que não é o próprio construto. Mais especialmente, um comportamento derivado de algo medido pelo teste, mas externo a ele, ou seja, nesse tipo de validade, há uma medida de algo externo ao teste. Estas peculiaridades podem se aplicar

a testes medindo o mesmo construto, construtos relacionados ou construtos diferentes avaliados ao mesmo tempo ou com um intervalo de tempo no qual houve uma intervenção.

Entre os critérios mais frequentemente empregados na validação dos testes, pode-se descrever o desempenho acadêmico, desempenho em treinamento especializado, desempenho profissional, diagnóstico psiquiátrico, diagnóstico subjetivo e outros instrumentos. Nem sempre é possível obter medidas de critério exatas, sendo, às vezes, ambíguas, dificultando a representação da variável que se deseja medir. Ainda a esse respeito, a evidência de validade baseada no critério pode ser estabelecida a partir de dois métodos diferentes, a saber, preditivo e concorrente. A validade preditiva consiste no que se pode esperar em termos de habilidade ou comportamento que o sujeito venha a desenvolver no futuro. Em outras palavras, indica a eficiência de um instrumento ao predizer algum resultado futuro. Vários métodos são utilizados para verificar a validade preditiva, no entanto, o que há de comum em todos eles é o uso de correlação do resultado dos testes e as condutas subseqüentes tomadas como critérios (Muñiz, 2004).

Já para Pasquali (2001), a validade concorrente diz respeito à correlação entre o resultado do teste e o critério, sem mensurar o tempo. Seu uso é importante para a descrição e diagnóstico de uma conduta que existe num dado momento, enquanto que a preditiva visa prever comportamentos. Vale destacar que a validade de um teste é expressa pelo coeficiente correspondente à correlação entre os escores obtidos e a medida de critério definida para a população a que se destina o teste. Assim, o coeficiente de validade, para ser estatisticamente significativo, deve ser suficientemente elevado, com os níveis de 0,01 e

0,05, embora haja alguma discordância entre os autores ao que se refere aos valores mais apropriados.

Quanto à validade de conteúdo, Anastasi e Urbina (2000), por sua vez, entende ser necessária a informação de especialistas no assunto tratado. Ela envolve a análise dos itens do teste no sentido de verificar se eles cobrem todo o domínio do construto (conhecimento). O conteúdo precisa, portanto, ser amplamente definido e, além disso, o estabelecimento da validade depende da relevância das respostas do teste para a área de comportamento em consideração, e não da aparente relevância do conteúdo dos itens.

Feita essas considerações sobre a importância dos parâmetros psicométricos para a utilização e comercialização de testes psicológicos, apontam-se, a seguir, algumas pesquisas já realizadas com o Teste Bender-SPG. No intuito de buscar evidências de validade desse novo sistema de correção, tendo em vista a finalidade específica desta pesquisa, em particular, que vislumbra a investigação de evidências de validade convergente-discriminante, bem como de evidências de validade de critério para o Bender-SPG. Além disso, observar-se à a precisão do instrumento.

PESQUISAS COM O BENDER-SPG

O Bender, muito utilizado no Brasil (Noronha *e cols*, 2002), esteve sempre presente, nas grades curriculares dos cursos de Psicologia das várias universidades ao longo dos últimos anos anteriores a 2003 (Pereira & Carelhos, 1995). Todavia, em função da proibição do uso de instrumentos que não atendem aos requisitos mínimos propostos pelo Conselho Federal de Psicologia, por meio da Resolução 02/2003, o instrumento deixou de ser aplicado e, provavelmente, ensinado por falta de estudos de validade e precisão com amostras brasileiras. Esses fatos justificam a relevância de pesquisas em âmbito nacional a respeito do Teste Gestáltico Visomotor de Bender, bem como de projetos de revisão.

A necessidade de que os testes se constituam como instrumentos confiáveis de avaliação, pautados nos princípios éticos e científicos, faz surgir novos trabalhos, a fim de validar o novo sistema de correção do Bender, propósito desta pesquisa, como já explicitado anteriormente, a partir do Sistema de Pontuação Gradual, (B-SPG). Entre estudos recentes, destaca-se o de Suehiro (2005), que teve por objetivo buscar evidências de validade de critério para o Bender-SPG, pela comparação de grupos divididos por dificuldades de aprendizagem (acentuada, média, leve, sem indícios) e pelo nível de escolaridade. Participaram de tal estudo 287 crianças, de ambos os sexos, entre 7 e 10 anos, de escolas públicas e particulares do estado de São Paulo. Quanto aos instrumentos, foram utilizados um questionário de identificação dos sujeitos, a Escala de Avaliação de Dificuldades na Escrita (ADAPE) e as nove figuras do Bender, que foram aplicadas coletivamente em uma única sessão. Os resultados apontaram diferenças significativas

entre as pontuações totais dos testes ADAPE e B-SPG relacionadas à idade, gênero e natureza jurídica da escola – pública ou particular –, bem como permitiram a identificação de evidências de validade de critério para o Bender, no que se refere às dificuldades de aprendizagem e aos diferentes níveis de escolaridade.

Por sua vez, o uso do B-SPG em crianças surdas foi objeto da investigação de Neri (2005), que as comparou com crianças ouvintes, com vistas a identificar se ocorre a mesma graduação associada à maturação visomotora. Participaram 139 crianças de ambos os sexos, divididas em dois grupos, sendo 19 com surdez severa ou profunda, com idades entre 9 e 11 anos e 120 ouvintes, de 6 a 11 anos. As crianças surdas eram de uma escola especializada do estado de São Paulo e as ouvintes foram selecionadas por sorteio em um Banco de Dados do Laboratório de Psicologia Educacional do Programa de Pós Graduação Stricto-Sensu em Psicologia da Universidade São Francisco (USF). Quanto aos materiais, foram utilizados um questionário de identificação de dados sobre a surdez, manual de aplicação e de correção do Teste de Bender, protocolos de registro das respostas das crianças surdas ao Bender e um retroprojetor. O instrumento foi aplicado às crianças surdas coletivamente em sala de aula, e com instruções mediante linguagem oral e de sinais.

Os resultados apontaram a não-sensibilidade da medida para captar o aspecto evolutivo da habilidade visomotora no grupo de crianças surdas, considerando que não foi observada a diminuição do número de erros conforme a idade aumentava. Também não foram observadas diferenças estatisticamente significativas quanto ao desempenho visomotor entre crianças surdas e ouvintes. Segundo a autora, os resultados podem ser atribuídos ao pequeno número de crianças surdas e a diferenças no procedimento de

aplicação do teste, com possível favorecimento desse grupo. O fato de elas dominarem a língua de sinais pode ter contribuído para a não-diferenciação entre o seu desempenho e o das crianças sem problemas de audição.

Já no que compreende a avaliação entre maturação percepto-motora e traços de personalidade, Santos (2006) utilizou o Teste de Bender (Sistema de Pontuação Gradual B-SPG) e a Escala de Traços de Personalidade para Crianças (ETPC), de Sisto (2004) com o objetivo de verificar se havia correlação entre maturação percepto-motora e traços de personalidade. Os instrumentos foram aplicados coletivamente em uma única sessão. Participaram dessa pesquisa 162 crianças, entre meninos e meninas, com idades variando entre 7 e 10 anos, de escolas públicas e particulares do estado de São Paulo. Na comparação entre médias de participantes dos dois sexos, não houve diferença estatisticamente significativa, porém os meninos obtiveram melhores resultados em relação às meninas. Verificou-se também que o Teste de Bender-SPG apresentou diferença (não significativa) entre alunos de escolas públicas e particulares, mas os alunos das escolas particulares tiveram um desempenho melhor que os das públicas. Foram encontradas diferenças significativas entre as idades dos participantes.

No item série, constatou-se que, conforme os alunos são aprovados de um ano para outro, e ficam mais velhos, sua pontuação nos protocolos nas figuras do Teste de B-SPG diminui. Esta diferença também foi significativa. A correlação entre a pontuação total dos protocolos do Teste de B-SPG com Traços de Personalidade foram significativas para todos os traços, exceto Sociabilidade. Nesta comparação, ao incluir resultado de Grupos Extremos, o traço de Sociabilidade também foi o único que não obteve correlação

significativa. Ao lado disso, no estudo de Sisto, Bueno & Rueda (2003) observou-se que quanto maior a intensidade do Traço de Personalidade Psicoticismo, maior a pontuação do critério de Distorção de Forma. Esse resultado pode indicar influência da personalidade na maturação percepto-motora em crianças. Evidentemente, muitos outros estudos serão, ou estão sendo desenvolvidos, o que permite valorizar cada vez mais o sistema de correção B-SPG em relação às suas qualidades psicométricas.

Bartholomeu (2006) buscou investigar evidências de validade entre o Teste Gestáltico de Bender avaliado pelo Sistema de Pontuação Gradual e o Desenho da Figura Humana (Sisto, 2005). Partiu-se da premissa de que os testes mencionados apresentariam em comum certos aspectos como o fato de ambos serem utilizados como preditores das capacidades intelectuais de crianças e envolverem aspectos motores. Buscaram-se, assim, relações entre a inteligência e a maturidade visomotora de crianças.

Os escores foram relacionados, separadamente, por sexo e idade. Além disso, os grupos extremos formados em razão da medida do DFH foram analisados em relação a possíveis diferenças nas pontuações do Bender. Participaram da pesquisa 244 alunos (50% meninos), com idades de 7 a 10 anos, média de oito anos ($DP=1,08$), que cursavam de primeira a quarta série do ensino fundamental de uma escola pública do interior do estado de São Paulo.

Foram evidenciadas correlações negativas e significativas entre os escores dos testes em ambos os sexos e nas idades, bem como diferenças entre os grupos extremos do DFH, em relação ao escore do teste de Bender, em todas as situações estudadas. Concluiu-se que o Teste Gestáltico foi um bom preditor das capacidades intelectuais de crianças, além de

diferenciar aquelas com um desenvolvimento intelectual acima da média das que apresentaram menor pontuação neste teste de inteligência.

Apesar de ser funcional e promissora a proposta do Sistema de Pontuação Gradual para análise do Teste de Bender, necessita, como qualquer outro sistema, ser mais bem estudado quanto às suas características psicométricas, com vistas a suprir as limitações encontradas nos demais sistemas existentes. Neste sentido, esse estudo poderá contribuir na implementação dessas características, ampliando as validades de evidências deste teste, bem como contribuir com mais um estudo no sentido de preencher a lacuna referente às questões da aprendizagem da leitura e da escrita em crianças de 6 a 10 anos do ensino público e privado.

Entende-se que muitos outros estudos virão e, da mesma forma, buscarão contribuir com dados relevantes na área de avaliação e no fomento de novas pesquisas para o setor educacional. Para tanto, os instrumentos utilizados devem ser apropriados e coerentes para a realidade vivenciada pelas crianças brasileiras e pautados em princípios científicos. A esse respeito, são importantes os estudos que se proponham a verificar se de fato os instrumentos que estão sendo empregados medem o que se propõem a medir, considerando as populações específicas para a qual foram construídos, e, quão bem o fazem. Com base nessas considerações, bem como nas demais apresentadas nos capítulos anteriores, este estudo foi proposto com os objetivos detalhados a seguir.

OBJETIVOS

Constituíram objetivos desta pesquisa:

- (1) Buscar evidências de validade convergente-discriminante por meio da correlação do B-SPG com os seguintes instrumentos: Escala de Avaliação de Dificuldade na Aprendizagem da Escrita (ADAPE), Teste de Cloze e o teste de Reconhecimento de Palavras;
- (2) Buscar evidências de validade de critério para o B-SPG pela comparação de grupos divididos por melhores e piores resultados nos instrumentos Cloze, ADAPE e Reconhecimento de Palavras;
- (3) Verificar a Precisão do B-SPG;
- (4) Estabelecer diferenças de médias dos instrumentos entre os participantes no que se refere às variáveis idade, sexo, série e tipos de escola.

O próximo capítulo tratará do Método, onde serão descritos os participantes, instrumentos utilizados e o procedimento de aplicação e correção dos protocolos.

CAPÍTULO III

MÉTODO

Participantes

Participaram do estudo 297 crianças, ambos os sexos, com idades entre 7 e 10 anos, sendo 144 do sexo feminino (48,5%) e 153 do sexo masculino (51,5%). Quanto à escolaridade, 18 pertenciam à 1ª série (6,1%); 24 à 2ª série (8,1%); 118 frequentavam a 3ª série (39,7%) e 137 estavam na 4ª série (46,5%). Do total, 81 estudavam em Escola Particular (27,3%), 216 em Escola Pública (72,7%).

Em relação à idade, 22 sujeitos tinham 7 anos (7,4%), 43 sujeitos, 8 anos (14,5%), 132, 9 anos (44,4%) e 100, 10 anos de idade (33,7%). A média de idades dos sujeitos era de 9,04 (DP=0,88).

Instrumentos

Os instrumentos utilizados para a realização do estudo foram os Testes Gestáltico Visomotor de Bender: Sistema de Pontuação Gradual (B-SPG), o Teste de Compreensão em Leitura (Cloze), a Escala de Avaliação de Dificuldade na Aprendizagem da Escrita (ADAPE) e o Teste de Reconhecimento de Palavras.

O TESTE GESTÁLTICO VISOMOTOR DE BENDER: SISTEMA DE PONTUAÇÃO GRADUAL (B-SPG) (Sisto, Noronha & Santos 2005)

O Bender-SPG é um teste que apresenta um sistema de correção com pontuação gradual. Surgiu em função da necessidade de implementar pesquisas que evidenciassem a validade e a precisão para muitos dos usos do Bender, assim como em razão da importância de se desenvolver novas propostas de sistemas de correção considerando, entre outras, possíveis limitações na interpretação das pontuações obtidas do modelo proposto pela pesquisadora Koppitz (1989), para investigação da percepção visomotora em crianças brasileiras.

Nessa perspectiva, Sisto, Noronha e Santos (2005) realizaram um amplo estudo para desenvolver um sistema de correção alternativo ao de Koppitz (1975), (o B-SPG), com o objetivo de utilizar a distorção de formas como critério de correção. Os itens de cada desenho são avaliados pela qualidade da reprodução de pontos, linhas retas, linhas curvas e ângulos, de acordo com a severidade dos desvios. Dessa forma, os autores selecionaram os critérios de correção do Sistema Koppitz (1975), que se mostraram mais adequados a partir da análise da variância (ANOVA) no que se refere à discriminação dos níveis de maturação visomotora, de idade em idade, dos 6 aos 10 anos.

Um dos principais focos desta proposta foi a manutenção do caráter evolutivo para o Bender-SPG, desenvolvido por Sisto, Noronha e Santos (2005), utilizando-se apenas do critério de Distorção da Forma, entendida como deformações que incluem desde alterações visíveis na configuração geral do desenho até figuras totalmente deformadas. De acordo com a gravidade do erro, pontua-se de 1 a 3. Em uma análise posterior, foram considerados

dois aspectos, a saber, as correlações entre as idades e os escores do Teste; as diferenças de média dos grupos, por idade, com a possibilidade de discriminação entre eles. Nessa perspectiva, ficou evidenciada a validade de construto referente à mudança desenvolvimental detectada.

Na correlação de Pearson entre idade (6 a 10 anos) e pontuação no B-SPG, encontrou-se um índice de correlação $r= 0,58$, significativo a $p\leq 0,000$. Esse dado indica que, conforme a idade aumentou, diminuiu a pontuação obtida no Teste de B-SPG, ou melhor, as distorções na reprodução das cópias diminuíram progressivamente com o aumento da idade. No aspecto relacionado às diferenças de médias entre os grupos de idades (6 a 10 anos) e à possibilidade de discriminação entre eles, utilizou-se a análise da variância e o teste de Tukey. Na análise da variância, consideraram-se as variáveis idade e sexo (4x2).

O resultado apontou a idade como fonte de maior variância, sendo a única que se mostrou estatisticamente significativa. Quanto ao sexo, os resultados não foram significativos, ou seja, não se mostrou como fonte de variância suficiente para produzir diferenças que não podem ser atribuídas ao acaso. Podem-se verificar, por meio dos resultados desse estudo, indicativos de evidências de validade de construto, em relação ao aspecto desenvolvimental. (Sisto, Noronha & Santos, 2005).

Já para a análise da dimensionalidade do teste e dos itens, utilizou-se o modelo de Rasch (1960) o qual supõe que o participante ao responder a um item do teste manifeste alguma quantidade de certa habilidade. *O funcionamento diferencial dos itens* (DIF) indica diversas propriedades estatísticas em razão de diferentes grupos, que em um teste válido,

fornecerá medidas idênticas para sujeitos ou grupos com níveis semelhantes da habilidade que o teste se propõe a medir. Os resultados evidenciaram apenas um fator que explicou 1,7 de 10 unidades de variância residual. Os itens escolhidos com saturação fatorial positiva foram a figura 1 e 5. Os com saturação fatorial negativa escolhidos foram as figuras 7a, 7b e 8. Esses itens positivos e negativos foram submetidos, separadamente, a uma nova análise pelo modelo do Rasch e correlacionados às medidas Rasch dos itens positivos e negativos de cada pessoa. O coeficiente de correlação de Pearson foi $r = 0,299$, o que indica que os dados possuem apenas uma dimensão.

Quanto às normas de interpretação dos resultados são fornecidas em relação a duas medidas. Uma delas é referente à pontuação bruta, produto da avaliação dos protocolos pela aplicação direta dos critérios; e a outra, pela medida de Rasch para os que tenham necessidade de uma informação dessa natureza.

O TESTE DE CLOZE NA AVALIAÇÃO DA COMPREENSÃO EM LEITURA (Santos, 2005)

O teste de Cloze constitui uma técnica que vem sendo usada por pesquisadores, professores e profissionais que avaliam a leitura e a linguagem, principalmente a partir dos anos 70 e consiste na seleção de um texto de aproximadamente 200 vocábulos, do qual, na proposta original de Taylor (1957), omite-se o quinto vocábulo como uma forma mais adequada para o diagnóstico da compreensão. Os examinandos devem preencher a lacuna com a palavra que julgarem ser a mais apropriada para a constituição de uma mensagem

coerente e compreensiva. Os escores são obtidos somando-se os números de lacunas preenchidas corretamente.

Esse teste tem se mostrado bastante eficaz, tanto do ponto de vista prático, tendo em vista a facilidade de elaboração, aplicação e correção, como do ponto de vista empírico, em função dos altos índices de correlação positiva de seus resultados com o desempenho acadêmico; isto é, alunos com maiores percentuais no teste apresentam melhores resultados nas médias das disciplinas (Manarini, 1996; Santos, 1990). Tradicionalmente, os significados dos escores do teste de Cloze seguem os parâmetros de Bormuth (1968) para analisar o desempenho dos sujeitos, ou seja, um percentual de acertos de até 44% do total de lacunas do texto caracteriza que o leitor teve pouco êxito na compreensão. Já um percentual de acertos entre 44% e 57% do texto mostra que a compreensão é suficiente, mas não a ideal, e aponta a necessidade de ajuda externa. Um nível de autonomia na compreensão do leitor seria um resultado acima de 57% de acertos.

Santos, Primi e Vendramini (2002) investigaram as propriedades psicométricas de um instrumento de avaliação da compreensão em leitura utilizando a técnica de Cloze. Participaram 612 estudantes ingressantes de quatro cursos de uma universidade particular, cuja tarefa era preencher as lacunas do texto com as palavras que julgassem mais apropriadas. Observou-se um bom ajuste ao modelo de dois parâmetros da Teoria de Resposta ao Item, sendo que a escala formada pela soma do número de lacunas corretamente completadas provou possuir consistência interna alta.

Os itens formaram dois grupos em razão da dificuldade: um grupo com itens muito fáceis e outro com itens muito difíceis. Parte da variabilidade dos índices de dificuldade

pôde ser explicada pela classe gramatical da palavra a ser descoberta ($r=0,62$, $p < 0,001$). Do maior ao menor nível de complexidade foram observados: adjetivos, verbos, advérbios, substantivos, pronome, preposição e artigo. Analisando-se os processos cognitivos que poderiam explicar as diferenças de dificuldade correlacionadas com as classes gramaticais e aliando-se à aplicação da Teoria de Resposta ao Item (TRI), discutiu-se o avanço das interpretações de proficiência em leitura referenciadas nos itens.

No sentido de avaliar a eficácia da técnica como instrumento de diagnóstico e de desenvolvimento da compreensão em leitura, Santos (2004) realizou um estudo em alunos da 5ª série. Foram aplicados pré e pós-testes e um programa de intervenção com uso do Cloze gradual em textos retirados de livros didáticos. A comparação entre eles apontou para uma diferença significativa, demonstrando que o Cloze pareceu adequado para o desenvolvimento da compreensão em leitura. A autora ressaltou que é um instrumento que tem se mostrado particularmente atraente para uso em sala de aula, pelo fato de ser de elaboração e correção bastante simples. Mas, alguns cuidados mínimos devem ser tomados, pois sua construção de forma indiscriminada pode acarretar sérios erros de avaliação.

Em razão dos poucos estudos de instrumentos desta natureza e de sua importância no auxílio em avaliação da aprendizagem, Gomes e Boruchovitch (2005) realizaram uma pesquisa que teve como objetivo analisar a eficácia do jogo Bingo Melhor Estudante e verificar as relações entre o potencial dos participantes no jogo, no Teste de Cloze e em uma Escala de Estratégias de Aprendizagem. Os sujeitos foram 29 alunos de 4ª série do ensino fundamental de uma escola pública do interior de São Paulo. Os resultados

evidenciaram correlações positivas e significativas entre o desempenho no jogo, a escala de estratégias de aprendizagem e o Cloze.

Num outro estudo, Santos, Rueda e Bartholomeu (2006) partindo-se do pressuposto de que a baixa compreensão em leitura é um dos indicadores das dificuldades de aprendizagem consideraram pertinente explorar sua relação com os aspectos emocionais manifestados pelas crianças que apresentam tais dificuldades. Nessa perspectiva, realizaram a pesquisa com o objetivo de comparar os grupos extremos de pontuações no Teste de Cloze, considerando-se os indicadores emocionais observados nos desenhos, conforme proposto por Koppitz (1968), buscando diferenciar problemas emocionais em crianças com e sem dificuldades de leitura, bem como eventuais diferenças relacionadas ao sexo.

Participaram da pesquisa 77 sujeitos da 4ª série do ensino fundamental de uma escola pública do interior de São Paulo com idade variando entre 9 a 13 anos. As crianças foram retiradas de uma amostra maior, selecionando-se 25% que apresentaram as melhores e as piores pontuações no Teste de Cloze. Dessa forma, o grupo 1 foi composto por crianças que possuíam até 8 acertos e o grupo 2, por crianças com 11 acertos ou mais. Os resultados foram analisados quantitativamente e qualitativamente em função dos objetivos estabelecidos. A média e os valores mínimos e máximos do grupo 1 foram baixos, indicando que o Teste de Cloze foi difícil para essas crianças. Foram evidenciados problemas na compreensão da leitura, haja vista, que em média acertaram menos que 50% dos pontos possíveis. Já para as crianças do grupo 2 os resultados revelaram uma melhor compreensão do texto, visto que a média de acertos ficou acima de 70% dos pontos possíveis. No entanto, os indicadores emocionais apontados por Koppitz, não diferenciaram

as crianças de ambos os grupos, ou seja, as crianças que apresentaram um pior desempenho no Cloze, não apresentaram mais indicadores emocionais do que as crianças que tiveram um melhor desempenho, e vice-versa.

Conforme apontam outros estudos, o Cloze pareceu ser uma técnica adequada para a avaliação da compreensão em leitura, distinguindo crianças com bom e mau desempenho em leitura (McKenna, 1979; Chance, 1975, 1985; Bensoussan, 1990; Bitar, 1989; Santos, 2004).

Para a realização desta pesquisa, foi selecionado o texto “A Princesa e o Fantasma” (Anexo 2) para ser utilizado, considerando que a população alvo do projeto constituiu-se de crianças na faixa de 7 a 10 anos. Este material foi elaborado por Santos (2005), que procurou manter os mesmos critérios de aplicação e correção da versão original de Taylor (1957). O material foi submetido a um estudo piloto com 314 alunos de ambos os sexos, de 2ª a 4ª séries, para verificar sua validade de critério.

Os resultados obtidos indicaram que o texto, com total de 103 palavras, das quais foram omitidas 15, mostrou-se adequado para utilização na amostra pretendida. A análise da variância indicou haver diferença estatisticamente significativa entre as séries [$F(3, 314) = 55,75; p = 0,0001$]. Essa diferença foi justificada pelo teste de *Tukey*, que separou os escores das crianças nos grupos correspondentes às séries freqüentadas.

A análise da consistência interna apresentou índices de precisão altos, de acordo com as recomendações do CFP (2003), acima de 0,60, pois o alfa de Cronbach foi de 0,83 para as crianças estudadas. Procedeu-se também à análise da consistência por série, sendo que se

obteve o índice de 0,85 para a segunda série, de 0,69 para a terceira série e de 0,72 para a quarta série.

Quanto ao critério de correção, deve-se atribuir um ponto para cada acerto, ou seja, para as palavras que foram escritas de forma idêntica à omitida. O número de acertos máximos no texto é de 15 palavras, correspondentes ao número de omissões.

ESCALA DE AVALIAÇÃO DE DIFICULDADE NA APRENDIZAGEM DA ESCRITA (ADAPE) (Sisto, 2001)

O ADAPE foi elaborado a partir da constatação das dificuldades mais comuns na escrita entre crianças de 1ª e 2ª séries. Objetiva avaliar a dificuldade de representação de fonemas, ou seja, de grafar letras e palavras por meio de um sistema lingüístico estruturado e aleatório. Para essa avaliação, foi elaborado um texto de 114 palavras; 60 delas apresentavam algum tipo de dificuldade classificada como encontro consonantal, dígrafo, sílaba composta e sílaba complexa; 54 palavras não apresentavam qualquer dificuldade. Cada uma delas foi considerada um item ou unidade de medida para efeitos de pesquisa. O instrumento consiste em um ditado chamado “Uma tarde no campo” (Anexo 3).

A primeira tarefa de Sisto (2001) para o desenvolvimento do texto do ditado foi verificar as dificuldades que mais apareciam na escrita entre crianças de primeiras e segundas séries, utilizando, para tanto, palavras usuais do material didático delas. Com esse levantamento, foi possível definir alguns critérios para a construção do texto, como palavras do cotidiano das crianças; em uma mesma palavra, poderiam aparecer mais de

uma dificuldade e pelo menos um terço das palavras deveriam ser trissílabas ou polissílabas.

A aplicação do ditado se faz a partir de uma palavra de cada vez e nenhuma delas deve ser repetida. Por isso, é necessária bastante atenção. Os critérios para a correção dos ditados baseiam-se na contagem dos erros de grafia, das ausências de palavras, acentuação errada, uso indevido de letras maiúsculas e minúsculas. A pontuação de cada criança corresponde à soma dos erros cometidos.

Quanto à validade, num primeiro momento, o objetivo centrava-se em saber se o instrumento em questão discriminava as crianças consideradas alfabetizadas das que estavam em processo de alfabetização. O autor identificou a existência de evidências de validade de critério, no estudo de 302 crianças de ambos os sexos, que freqüentavam a 1ª série do ensino fundamental no município de Campinas. Destas, 94% fizeram a pré-escola e iniciaram a alfabetização na primeira série, com idade de 7 anos. O percentil de erros foi baseado nos resultados obtidos. As crianças que estavam sendo alfabetizadas pela forma tradicional de ensino por professoras reconhecidas como boas alfabetizadoras foram consideradas como crianças-critério e admitidas como modelo para verificar o alcance da aprendizagem. Elas apresentaram até 29 erros.

Na avaliação realizada, o autor concluiu que 75% das suas crianças estavam alfabetizadas ou muito próximas disso. Por sua vez, as crianças consideradas alfabetizadas ou quase pelas professoras das outras classes apresentaram até 38 erros. Um número bastante grande de crianças desse 2º grupo obteve pontuação entre 30-38, levando à conclusão de que aproximadamente 30% dessas crianças poderiam ser consideradas

alfabetizadas. De fato, a pontuação de 29 erros está entre o percentil 20-25, e a pontuação de 38 erros está entre o percentil 25-30, sugerindo uma proximidade de critério de avaliação entre as diferentes professoras.

Após essa análise inicial, uma segunda coleta de dados foi realizada com o objetivo de se obter a precisão do instrumento e estabelecer normas para a comparação do aluno com seus pares, como também formular critérios para a classificação de níveis de dificuldades de aprendizagem da escrita.

O mesmo ditado foi aplicado a outras 770 crianças de primeiras e segundas séries, de ambos os sexos e idade entre 7 e 13 anos, estudantes do ensino público do município de Campinas. Nesse grupo, foram incluídas as crianças-critério analisadas no estudo de validação. Assim, formaram-se três grupos: crianças-critério, crianças de 1ª séries comuns e crianças de 2ª séries.

A análise da consistência interna do instrumento foi realizada por meio do método das metades. Os dados de precisão do instrumento foram suficientemente altos, tanto para os três grupos separadamente quanto para todas as crianças juntas. Pôde-se verificar que, de uma maneira geral, a porcentagem de erros por palavras diferenciou os grupos claramente, ou seja, as crianças-critério (alfabetizadas pelas professoras) obtiveram menos erros em cada palavra, as de segunda série mais erros e as de 1ª série superaram o número de erros, em comparação com os dois primeiros grupos. Quanto aos coeficientes de Spearman-Brown para as crianças-critério de primeiras e segundas séries, variaram de 0,87 a 0,97 (Sisto, 2001).

Com o propósito de desenvolver normas e a classificação para detectar crianças com possíveis dificuldades de aprendizagem na escrita, Sisto (2001), levou em conta os critérios de distanciamento do ponto de aquisição de escrita descrito pelas professoras e mostrado pelo instrumento nas crianças-critério e de primeira série. Considerando que, no estado de São Paulo, o sistema de ensino não reprova no mesmo ciclo, percebeu-se a necessidade de mais um ano para que as crianças completassem a aprendizagem e dominassem a escrita. Os critérios de Classificação de Dificuldade de Aprendizagem na Escrita do ADAPE para as 2ª e 3ª séries estão registrados no Anexo 3. Esses, não serão utilizados nesta pesquisa, e sim, a pontuação bruta apresentada na escala.

Utilizando-se dos critérios de classificação Bazi e Sisto (2006) buscaram relacionar as dificuldades de aprendizagem na escrita às emoções alegria e tristeza, medo e coragem. Neste estudo, as dificuldades de aprendizagem foram compreendidas como relacionadas a um grupo heterogêneo de transtornos e a dificuldades significativas de aquisição e de uso da recepção, da fala, da leitura, da escrita, do raciocínio e das habilidades matemáticas, e também por se referirem as dificuldades intrínsecas, que podem ocorrer ao longo da vida. Partiram-se das hipóteses que crianças com altos níveis de dificuldades de aprendizagem na escrita apresentariam baixos níveis de alegria, ocorrendo o inverso com as crianças mais tristes, e também que apresentariam maiores níveis de medo, ocorrendo o contrário com os sujeitos com altos níveis na pontuação de coragem.

Participaram da pesquisa 602 sujeitos, sendo 293 pertencentes a 2ª série do ensino fundamental e 309 pertencentes à 3ª série. A faixa etária era de 8 e 10 anos. Sisto (2001a) relata ser possível falar em dificuldades de aprendizagem, somente a partir do segundo

semestre do segundo ano, pois garante mais um ano à alfabetização. Nesse sentido, segundo os autores, justifica-se a escolha para aplicação somente em alunos das 2ª e 3ª séries.

A avaliação das dificuldades na escrita foi realizada por meio da ADAPE (Sisto, 2001^a). As medidas das emoções alegria e tristeza, medo e coragem, foram obtidas pelo instrumento Autopercepção de sentimentos (Sisto & Bazi, 1999). A pontuação se baseia em atribuir um ponto para cada afirmação do sujeito. O instrumento contém 32 itens, sendo 8 para cada emoção (os escores podem variar de 0 a 8 pontos para cada emoção). Os instrumentos foram aplicados coletivamente.

Os resultados foram apresentados por série. Para a 2ª série, a análise da relação das variáveis medo e coragem com a variável dificuldade de aprendizagem na escrita, não forneceram resultados significativos [(F=0,67); $p=0,57$], [(F=1,27); $p=0,28$]. Dessa maneira, a hipótese de que crianças com maiores níveis de dificuldade de aprendizagem teriam mais medo não foi comprovado, como também não se confirmou a hipótese de que as crianças com alto nível de dificuldade de aprendizagem, teriam menos coragem.

Quanto à 3ª série, à medida que aumentou o nível de dificuldade de aprendizagem em escrita, houve diminuição do nível de alegria. A hipótese de que as crianças com alto nível de dificuldade de aprendizagem pudessem apresentar baixos níveis de alegria pôde ser confirmada pela análise de variância. Já a hipótese de que crianças com maiores níveis de dificuldades de aprendizagem pudessem apresentar maiores níveis de tristeza, não pôde ser comprovada.

A análise da relação das variáveis medo e coragem com a variável dificuldade de aprendizagem na escrita não forneceu resultados significativos [(F=0,94); $p=0,42$],

[(F=0,48); $p=0,70$], respectivamente. Assim, a hipótese de que crianças com maiores níveis de dificuldade de aprendizagem pudessem apresentar maiores níveis de medo não pode ser comprovada, como também não foi confirmada a hipótese de que crianças com alto nível de dificuldades de aprendizagem, teriam baixos níveis de coragem.

Em outro estudo Sisto e Rueda (2006) optaram pela utilização dos critérios de Machover para avaliar as características emocionais das crianças. Dessa forma, o objetivo da pesquisa foi identificar quais as características emocionais mais frequentes em crianças da 2ª série, assim como verificar possíveis diferenças emocionais em dois grupos, um com grande dificuldade de aprendizagem na escrita e outro sem dificuldades. A pesquisa foi realizada em 81 crianças da 2ª série de uma escola municipal do interior do estado de São Paulo com idade de 8 e 9 anos. Utilizaram para tanto, a Escala de Avaliação de Dificuldades na Aprendizagem da Escrita (Sisto, 2001) e o Teste do Desenho da Figura Humana que propõe avaliar os aspectos emocionais em crianças. Foram avaliados 12 itens: tamanho em relação à folha, localização na folha, tipo de linha, consistência do traçado, predominância de linhas retas, curvas ou ângulos, correções e retoques, sombreamento ou borrão, valor atribuído às partes do desenho, estereotipia, movimento, transparências e perspectiva.

Os resultados sugeriram que houve bastante dispersão das pontuações do total de erros por palavras em relação à média, de tal modo que as crianças obtiveram escores variados em relação aos erros na escrita. Procurou-se verificar até que ponto as crianças diferiam em termos de erros na escrita em razão do número de indicadores emocionais, por

meio da prova ANOVA. O resultado dessa análise indicou que as diferenças das médias não podem ser atribuídas ao acaso [(F=2,79); $p=0,004$].

As provas de *Tukey* mostraram que a diferença encontrada é justificada pela diferença do total de erros na escrita existente entre os grupos de crianças que evidenciaram 5 e 6 indicadores emocionais em seus desenhos das que apresentaram 8 indicadores, sendo que as crianças com 7 indicadores não se diferenciaram nem de um grupo nem do outro. Ao lado disso, verificou-se que o sexo masculino e o grupo total, independentemente do sexo, forneceram correlações positivas e significativas entre o total de erros por palavras e a soma de indicadores emocionais, enquanto que o sexo feminino não manteve correlações com as variáveis. Nesse sentido, os erros nas escritas das crianças parecem estar associados aos indicadores emocionais somente para meninos, revelando que quanto mais problemas emocionais esses evidenciam, maior a dificuldades de aprendizagem na escrita.

Considerando os resultados desta pesquisa, verificou-se que os grupos de crianças com menos indicadores emocionais nos desenhos apresentou menores médias de erros na escrita, em detrimento das com mais problemas emocionais. Além do que, os erros na escrita apareceram vinculados aos problemas emocionais dos sujeitos estudados. Esses dados sugerem que o agravamento dos problemas emocionais das crianças está associado a maiores dificuldades de aprendizagem na escrita. Várias outras pesquisas foram realizadas utilizando-se desse instrumento para investigar a compreensão da leitura de crianças com dificuldades de aprendizagem na escrita, dentre as quais se destaca: Bispo (2000), Zucolotto e Sisto (2002), Carneiro, Martinelli e Sisto (2003), Sisto e Fernandes (2004) e Pacheco e Sisto (2005).

TESTE DO RECONHECIMENTO DE PALAVRAS (Sisto, 2002)

No trabalho de Sisto e Kingeski (2002), as 221 palavras que compunham o instrumento para a avaliação do reconhecimento de palavras foram escolhidas pelo sorteio das apostilas e dos livros utilizados pelos próprios alunos no seu cotidiano escolar. Elas foram classificadas em palavras sem dificuldades e com dificuldades de escrita. Estas últimas poderiam apresentar encontros consonantais, dígrafos, sílabas compostas, sílabas complexas, com uma ou mais dificuldades.

Assim, o instrumento tomou a forma de um teste de múltipla escolha composto por 221 itens, distribuído em 7 páginas. Cada item possuía três alternativas de resposta, das quais somente uma era a correta. Apesar das características reveladas por este estudo quanto ao desempenho dos estudantes das séries iniciais do ensino fundamental no reconhecimento de palavras (por exemplo, a existência de diferenças de tempo, variando conforme a série; a ocorrência de diferentes níveis de desempenho, variando conforme o gênero; a ocorrência de diferentes níveis de desempenho variando conforme a série), algumas lacunas ficaram evidentes. Um novo estudo foi feito visando reduzir o número de palavras na composição do instrumento. Optou-se por 55 palavras ao invés das 221. Quanto ao critério de correção, a cada acerto atribuí-se um ponto, sendo possível atingir o escore máximo de 55 pontos (Anexo 4). Foram sugeridos estudos para transformar o material em uma escala de avaliação, na qual pudessem ser aferidas a consistência interna, a precisão e a validade.

Kingeski e Sisto (2004) realizaram outro estudo para avaliar o reconhecimento de palavras em 338 crianças, ambos os sexos, com idades entre 6 e 12 anos, que freqüentavam

da 1ª à 4ª série do ensino fundamental. Nesse sentido, utilizaram o instrumento Reconhecimento de Palavras com o objetivo de diferenciar as séries iniciais do ensino fundamental, de acordo com a classificação de Sisto (2001). O resultado demonstrou que os alunos do ensino fundamental utilizaram menos tempo para responder ao instrumento. Quanto à média de acertos, o gênero feminino apresentou-se superior em relação ao masculino. $[F(3, 338) = 92,270; p = 0,000]$ e $[F(3, 338) = 89,577; p = 0,000]$. A análise da variância entre as médias indicou uma diferença estatisticamente significativa entre as séries ($p < 0,001$), que foi verificada pelo teste de Tukey, que por sua vez, separou os escores das crianças nos grupos correspondentes às séries frequentadas, atribuindo ao instrumento a validade de critério.

Procedimento

Inicialmente foi feito contato com as escolas da rede pública e particular de uma cidade do interior do estado de São Paulo, cujos responsáveis autorizaram as coletas de dados. Os instrumentos foram aplicados coletivamente em sala de aula por uma equipe distribuída, com dois aplicadores por sala; um deles dava as instruções, e o outro auxiliava os sujeitos individualmente, quando necessário. A aplicação ocorreu após os pais terem concordado e autorizado a participação das crianças, assinando um Termo de Consentimento Livre e Esclarecido (Anexo 6), bem como após a aprovação do projeto pelo Comitê de Ética em Pesquisa da Universidade São Francisco, de acordo com a Resolução 196/1996 do Conselho Nacional de Pesquisa (CONEPE).

Para a aplicação dos testes, foram seguidas as orientações contidas nos manuais de cada instrumento. A pesquisadora treinou duas estagiárias para auxiliar na coleta dos dados e correção dos testes. Todos os instrumentos, Reconhecimento de Palavras, Cloze, Bender-SPG e o ADAPE foram aplicados no mesmo dia. Vale ressaltar que pelo fato da coleta ter sido realizada em dois tipos de instituições, a saber, pública e particular, foram realizados vários encontros, em horários alternados conforme a programação de cada instituição, para que fosse atingida toda a demanda. Mais especialmente, houve quatro visitas às escolas particulares e 10 às escolas públicas.

Ficou estabelecido que a aplicação ocorreria nas salas de aula, em grupos de no máximo 25 crianças. Teve-se a preocupação de orientar os participantes de que iriam realizar uma atividade diferente daquelas em que eles estavam acostumados a desenvolver em sala de aula. Esse procedimento foi utilizado no intuito de deixá-los mais à vontade e para a redução da ansiedade, visível naquele momento.

Um fato interessante ocorreu em alguns grupos de alunos da escola pública: - a curiosidade pelo retroprojeter. Eles vinham à frente e perguntavam o nome do instrumento e solicitavam que os deixassem manipulá-lo. Somente após satisfazer a curiosidade deles é que se dava início às atividades.

Quanto à seqüência na aplicação dos instrumentos, iniciou-se pelo Teste de Reconhecimento de Palavras. Em seguida, era aplicado o Teste de Compreensão em leitura (Cloze), O Bender-SPG e finalmente o ADAPE. Este foi deixado por último por envolver a habilidade na escrita e em razão da dificuldade que poderia gerar.

O material foi entregue individualmente para cada criança, que contavam de folhas de respostas do aluno, com uma lista de 55 palavras, escritas de três formas diferentes, sendo que o mesmo deveria marcar a palavra que considerava correta. Este material inicialmente aplicado era referente ao Teste Reconhecimento de Palavras (vide Anexo 4). Havia mais 2 folhas em branco, de papel sulfite para a reprodução das figuras de Teste Bender-SPG e para o ditado sobre o texto “Uma Tarde no Campo” do ADAPE.

Logo após a aplicação do Teste de Reconhecimento de Palavras, que durou em torno de 20 minutos, iniciou-se a orientação para o Teste de Compreensão em Leitura (Cloze). Este teste era composto do texto “A Princesa e o Fantasma,” sendo que os alunos deveriam realizar uma leitura silenciosa do material e a seguir, preencher os espaços com as palavras que consideravam adequadas e que dessem sentido ao texto. Vale lembrar, que era necessário preencher as lacunas em branco, com as respectivas palavras, de acordo com o tamanho da lacuna. A duração desse teste foi de aproximadamente quinze minutos.

O Outro instrumento realizado pelos participantes foi o Teste Bender-SPG. Para a aplicação do Bender, as figuras de A a 8 que o compõem, foram apresentadas uma a uma por meio de um retroprojeter. Quanto ao material, foi utilizada para a aplicação deste instrumento, uma folha de papel sulfite em branco, os modelos das figuras do Bender em transparência, e um lápis preto. A elaboração dos desenhos não prevê o uso de borrachas ou qualquer outro tipo de material que sirva de suporte mecânico.

Após execução da tarefa, que durou em torno de 20 minutos, deu-se início à aplicação do Teste ADAPE que se compõe de um ditado “Uma tarde no Campo”. Ao final dessa atividade que durou em torno de quinze minutos, as folhas foram recolhidas e

identificadas corretamente. Nenhuma marcação poderia ser feita no protocolo do Bender ou qualquer outro instrumento a não ser as respostas solicitadas de acordo com as instruções recebidas. A aplicação total dos instrumentos em cada grupo de alunos teve uma duração de uma hora e dez minutos.

O processo de correção iniciou-se logo após a coleta de todos os dados, a qual teve uma duração aproximada de 4 meses, divididos entre os dois bimestre finais do ano letivo de 2004 e início do bimestre de 2005. Para facilitar a correção, inicialmente separaram-se os instrumentos e a correção foi realizada pelas estagiárias e revisada pela pesquisadora. Os protocolos estavam previamente numerados e à medida que eles eram corrigidos, os dados foram transportados para o programa Excel, sendo que ao final, os resultados eram conferidos e transferidos para o programa SPSS, para futura análise. Os protocolos que estavam incompletos foram descartados, totalizando 50, o que justificou a grande desproporção entre o número de participantes de cada escola.

CAPÍTULO IV

RESULTADOS E DISCUSSÃO

Os resultados serão discutidos de acordo com os objetivos propostos para este estudo. Inicialmente serão apresentadas as estatísticas descritivas dos instrumentos. Em seguida realizar-se-a a análise da correlação do B-SPG com a Escala de Avaliação de Dificuldade na Aprendizagem da Escrita (ADAPE), os Teste de Cloze e o Teste de Reconhecimento de Palavras.

Num momento posterior, será apresentada a comparação de grupos divididos por melhores e piores resultados (grupos extremos) nos instrumentos Cloze, ADAPE e Reconhecimento de Palavras, bem como verificada a precisão do B-SPG. Por fim, serão discutidos os resultados encontrados nas diferenças de médias dos instrumentos entre os participantes no que se refere às variáveis sexo, idade, série e tipos de escola.

Estatísticas descritivas dos instrumentos

Nas Tabelas 1, 2 e 3, podem ser encontrados os dados relativos às estatísticas descritivas dos instrumentos aplicados.

Tabela 1 - Estatísticas descritivas dos instrumentos por série

Escolaridade	Teste	N	Mínimo	Máximo	Média	D P
1º série	Bender-SPG	18	2	18	8,83	3,944
	Rec. Palavra	18	10	49	29,56	11,089
	Cloze	18	0	15	7,28	4,084
	ADAPE	18	8	100	46,56	27,838
2º série	Bender-SPG	24	2	16	8,79	3,538
	Rec. Palavra	24	14	51	32,50	11,534
	Cloze	24	0	14	8,42	4,085
	ADAPE	24	0	114	36,33	26,023
3º série	Bender-SPG	118	0	20	7,21	4,078
	Rec. Palavra	118	0	55	37,93	11,268
	Cloze	118	0	15	9,42	3,815
	ADAPE	118	1	114	32,19	25,834
4º série	Bender-SPG	137	0	14	7,09	3,260
	Rec. Palavra	137	11	55	41,72	9,722
	Cloze	137	0	15	10,64	3,719
	ADAPE	137	0	114	27,50	21,545

As estatísticas descritivas dos instrumentos foram organizadas por série, tipo de escola e por idade. No que diz respeito aos resultados encontrados nos indicadores de erros ou acertos dos itens dos instrumentos por série, o mínimo de erros no Bender-SPG foi 0 para as 3ª e 4ª séries e 2 erros para os estudantes de 1ª e 2ª séries, sendo que a 3ª série obteve a maior média de erros (20 pontos). Vale destacar que neste instrumento pontuam-se os erros cometidos pelos sujeitos, e que a média desse grupo foi de 7,38 erros. No teste de Reconhecimento de Palavras, por outro lado, pontuam-se os acertos. O mínimo de acertos

foi obtido pelos estudantes da 1ª série (10) e o máximo pelos de 3ª e 4ª séries (55). No Cloze, o mínimo foi de 0 e o máximo de 15 acertos, para as 1ª, 3ª e 4ª séries. Quanto ao ADAPE, são pontuados os erros cometidos; nesse caso, o mínimo de erros foi 0, para as 2ª e 4ª séries e o máximo, 114 para as 2ª, 3ª e 4ª séries, fato que causou estranheza, sendo que na 1ª série, o máximo foi de 100 neste instrumento. Vale salientar que alunos da 1ª série ainda estão sendo alfabetizados.

Tabela 2 - Estatísticas descritivas dos instrumentos por tipo de escola

Instituição	Teste	N	Mínimo	Máximo	Média	D P
Escola Particular	Bender-SPG	81	0	17	7,38	3,872
	Rec.Palavras	81	16	55	43,90	8,995
	Cloze	81	2	15	11,73	2,855
	ADAPE	81	0	114	22,42	18,325
Escola Pública	Bender-SPG	216	0	20	7,38	3,638
	Rec.Palavras	216	0	55	36,79	11,238
	Cloze	216	0	15	9,04	4,005
	ADAPE	216	0	114	34,54	25,638

Quanto ao tipo de instituição, verificou-se que alunos das escolas particulares e públicas obtiveram a mesma média de erros no Teste de Bender-SPG ($M=7,38$). No entanto, no que se refere à pontuação máxima obtida, houve diferença entre as escolas, de tal modo que na escola particular obteve-se 17, e na pública, 20. Vale lembrar que o máximo de erros permitido no instrumento é de 21. No teste de Reconhecimento de Palavras o mínimo de acertos foi 16, apontando uma leve superioridade no desempenho do teste em estudantes de escola particular.

Os resultados evidenciaram ainda, uma tendência de pontuação mais baixa relacionada às dificuldades de aprendizagem da escrita (ADAPE). Nesse sentido, considerando que a pontuação máxima possível era de 114 pontos e que o instrumento utilizado baseia-se nos erros obtidos pelas crianças, verificou que os participantes da escola particular apresentaram menos erros (22,42) que os da escola pública (34,54). Esses resultados estão em consonância com o trabalho de (Machado, 1978; Suehiro, 2005; Santos, 2005, entre outros) já que os dados da autora revelaram superioridade na média dos participantes de escolas particulares.

Tabela 3 - Estatísticas descritivas dos instrumentos por idade

idade	Teste	N	Mínimo	Máximo	Média	D P
7	Bender-SPG	22	2	18	9,14	3,745
	Rec.Palavra	22	10	49	30,05	10,764
	Cloze	22	2	15	8,00	3,546
	ADAPE	22	8	100	42,36	25,550
8	Bender-SPG	43	2	1	8,86	4,057
	Rec.Palavra	43	0	55	35,51	12,115
	Cloze	43	0	15	9,37	4,494
	ADAPE	43	0	114	31,35	24,686
9	Bender-SPG	132	0	20	7,06	3,673
	Rec.Palavra	132	12	55	39,30	10,556
	Cloze	132	0	15	9,86	3,903
	ADAPE	132	1	114	30,30	24,832
10	Bender-SPG	100	0	14	6,79	3,304
	Rec.Palavra	100	11	55	41,27	10,354
	Cloze	100	1	15	10,22	3,653
	ADAPE	100	0	114	29,97	23,301

Na tabela que trata das estatísticas descritivas dos instrumentos por idade percebeu-se uma diminuição na média de erros no Teste de Bender-SPG, à medida que idade avançou. O mínimo de erros apresentados foi 0 e o máximo 20 em crianças de 9 anos. Os resultados são congruentes com os estudos de (Néri, 2005; Santos, 2005; Suehiro, 2005, entre outros), nas quais as crianças mais velhas obtiveram melhores desempenhos que as mais novas por possuírem mais domínio da motricidade gráfica (Machado, 1978; Mazzeschi & Lis, 1999; entre outros).

Correlação de *Pearson* entre os instrumentos, idade e escolaridade

Para se verificar a hipótese de que os construtos medidos pelos instrumentos utilizados estavam relacionados, recorreu-se à prova de Correlação de *Pearson*. Seguem as correlações entre os instrumentos e os respectivos níveis de significância. A Tabela 4 apresenta os coeficientes de correlação entre os instrumentos utilizados na pesquisa.

Tabela 4 - Correlação de *Pearson* entre o Teste Bender-SPG e os instrumentos Reconhecimento de Palavras, Cloze e ADAPE

		Bender- SPG	Reconhecimento de Palavras	Cloze
Bender	<i>r</i>			
	<i>p</i>			
Reconhecimento de Palavras	<i>r</i>	-0,29		
	<i>p</i>	0,000		
Cloze	<i>r</i>	-0,21	0,60	
	<i>p</i>	0,000	0,000	
ADAPE	<i>r</i>	0,31	-0,72	-0,55
	<i>p</i>	0,000	0,000	0,000

Os resultados evidenciaram que todas as correlações foram significativas no nível de 0,001. Como esperado, o Bender-SPG apresentou correlação positiva com o ADAPE ($r=0,31$; $p=0,000$) e correlações negativas com os demais. Com o teste de Cloze a correlação foi de $r=-0,21$ ($p=0,000$) e com o Reconhecimento de Palavras, de $r=-0,29$ ($p=0,000$). Os resultados confirmam que os construtos estudados estão relacionados, já que a análise entre os instrumentos indicou o quanto eles têm em comum. Em outra medida, não se pode afirmar que tratem de um único domínio teórico, tendo em vista que os coeficientes de correlação foram baixos. Nesse sentido, as correlações variaram de 0,21 a 0,31 (todas significativas), sugerindo que a maturidade percepto-motora está relacionada com compreensão de leitura e com a escrita. Em síntese, há evidências de validade de construto convergente-discriminante (Anastasi & Urbina, 2000). Em acréscimo, embora não seja objetivo desse estudo, a análise da correlação entre os instrumentos de compreensão em leitura (Cloze) e escrita (ADAPE e Reconhecimento de Palavras) foram de moderados a altos; reafirmando a comunalidade dos construtos.

Os achados estão em consonância com os de outros estudos já citados. O trabalho de Nielsen e Sapp (1991) realizado com 153 crianças norte-americanas entre 6 e 12 anos indicou que o teste de Bender tem sensibilidade suficiente para predizer habilidades de leitura em crianças nessa faixa etária. Confirmando esse atributo do Bender, segue a pesquisa documental realizada por Lesiack (1984), que revisou 32 pesquisas nas quais o Bender foi utilizado como medida de habilidade visomotora, relacionando-a com o desempenho em leitura. Este estudo revelou que os resultados do Bender mostraram evidências de validade concorrente com outras medidas de leitura. Ao lado disso, os

escores do Bender discriminaram claramente os bons e os maus leitores. Como conclusão a autora destacou que as pesquisas revisadas demonstraram que o Bender, pareceu ser preditivo do desenvolvimento da leitura.

Em que pese à escrita, pesquisas revelaram a correlação da maturação e do nível intelectual como um fator que influencia o seu desenvolvimento. A literatura aponta que as crianças mais velhas apresentam melhor desempenho do que as crianças mais novas em um mesmo nível escolar, considerando que, com o passar do tempo, a tendência é que haja melhor domínio da motricidade gráfica (Machado, 1978; Mazzeschi & Lis, 1999; entre outros).

A fim de investigar a relação entre os construtos e as variáveis idade e escolaridade, realizou-se a correlação de Pearson. Os dados podem ser melhores visualizados na Tabela 5 a seguir. Os coeficientes foram significativos, embora baixos. A exceção é feita para o ADAPE com idade, uma vez que a correlação não foi significativa. Dentre os coeficientes significativos, destaca-se a relação entre escolaridade e Reconhecimento de Palavras, uma vez que foi a maior correlação encontrada. Quanto ao Bender-SPG, a correlação com a idade mostrou-se maior que com a escolaridade, o que reafirma os estudos de Sisto, Noronha e Santos (2005).

Tabela 5 - Correlação de *Pearson* entre os Testes Bender, Reconhecimento de Palavras, Cloze e ADAPE e as variáveis Idade e Escolaridade (N= 297)

		Bender Total	Reconhecimento de Palavras	Cloze	ADAPE
Idade	<i>r</i>	-0,207	0,267	0,137	-0,099
	<i>p</i>	0,000	0,000	0,018	0,089
Escolaridade	<i>r</i>	-0,135	0,318	0,246	-0,195
	<i>p</i>	0,020	0,000	0,000	0,001

Estes resultados estão em consonância com os estudos de Bandeira e Hutz (1994), que investigaram o grau de predição do rendimento escolar na primeira série por meio da aplicação dos testes Desenho da Figura Humana (DFH), Bender e Raven. Os resultados apontaram para correlações significativas entre os três testes e o rendimento, embora apenas o Bender e os itens evolutivos do DFH tenham contribuído para explicar a variância do rendimento escolar. Também nessa linha, Suehiro (2005) realizou um estudo com o objetivo de buscar evidências de validade de critério para o Bender-SPG, pela comparação de grupos divididos por dificuldades de aprendizagem (acentuada, média, leve e sem indícios) e pelo nível de escolaridade. Participaram da pesquisa 287 crianças de ambos os sexos, entre 7 e 10 anos, de escolas públicas e particulares do estado de São Paulo. Os resultados permitiram a identificação de evidências de validade de critério para o Bender, no que se refere às dificuldades de aprendizagem e aos diferentes níveis de escolaridade.

Análise de Grupos Extremos

Tabela 6 - Grupos extremos do Bender em comparação com o Teste Reconhecimento de Palavras

Reconhecimento de Palavras		N	Média	DP	<i>t</i>	<i>p</i>
Bender Total	Grupo 1	75	8,81	4,16	4,13	0,000
	Grupo 3	74	6,26	3,32		

Para a análise dos grupos extremos, estes foram organizados em razão dos melhores e piores resultados do B-SPG. Utilizou-se, para tanto, a divisão por quartil (grupo 1 até 25 e grupo 3 após 75). Os resultados apresentados na Tabela 6 evidenciam que, em relação ao teste Reconhecimento de Palavras, os grupos do Bender diferenciaram-se significativamente ($t=4,13$; $p=0,000$).

Tabela 7 - Grupos extremos do Bender em comparação com o Teste de Compreensão em Leitura (Cloze)

Cloze		N	Média	DP	<i>t</i>	<i>p</i>
Bender Total	Grupo 1	74	8,73	3,79	1,87	0,06
	Grupo 3	86	7,65	3,47		

Para os grupos extremos do Bender, em comparação com o teste de Cloze, os resultados revelaram que foram marginalmente significativos, ($t=1,87$; $p=0,006$), tal como apresentados na Tabela 7.

Tabela 8 - Grupos extremos do Bender em comparação com o Teste ADAPE

	ADAPE	N	Média	DP	<i>t</i>	<i>P</i>
Bender Total	Grupo 1	76	6,07	3,14	-5,16	0,000
	Grupo 3	75	8,97	3,75		

A Tabela 8 apresenta a análise dos grupos extremos do Bender-SPG em relação ao ADAPE. Os resultados evidenciaram diferenças significativas entre os grupos extremos ($t=-5,16$; $p= 0,000$) e reforçam o pressuposto de que o desenvolvimento percepto-motor é requisito para o desenvolvimento de habilidades de leitura e escrita, considerando-se que o Teste de Bender tem sido revelado útil para o diagnóstico diferencial de crianças com problemas de aprendizagem, tal como preconizado por (Sisto, Noronha & Santos, 2005). Em síntese, pode-se ressaltar que o B-SPG foi capaz de discriminar os melhores e piores resultados em compreensão de leitura, no reconhecimento de palavras e no desempenho da escrita. Tais achados permitem indicar evidências de validade de critério do B-SPG.

Dando continuidade aos objetivos da pesquisa, verificou-se a precisão do Bender-SPG. Esta foi investigada por meio da verificação do Alfa de Cronbach. Constatou-se que o instrumento é preciso ($\alpha=0,765$), de acordo com Cronbach (1998). No que se refere à

contribuição de cada item para a precisão da escala, é possível observar que nenhum deles, caso fosse retirado, aumentaria substancialmente a precisão como pode ser verificado na Tabela 9. Considerando que o sistema de correção possui apenas dez itens, o coeficiente encontrado revela-se excelente.

Precisão do Bender-SPG

Tabela 9 - Índices de Precisão do Bender-SPG por Alfa de Cronbach

	Média por exclusão de item	Variância por exclusão de item	Correlação item-total	Alfa do item
FIG.A	14,153	52,048	0,341	0,732
FIG.1	14,302	51,358	0,367	0,729
FIG.2	13,823	52,460	0,319	0,734
FIG.3	13,748	49,165	0,559	0,714
FIG.4	14,023	49,149	0,573	0,714
FIG.5	13,972	49,016	0,563	0,713
FIG.6	13,411	48,306	0,539	0,711
FIG.7_A	14,034	48,217	0,599	0,708
FIG.7_B	13,938	49,525	0,590	0,715
FIG.8	14,312	50,038	0,528	0,719
Bender_T	7,398	13,865	0,995	0,765

Diferenças de médias – variáveis, sexo, série, idade e tipo de escola

Comparação de médias por sexo

A fim de verificar as diferenças de médias entre as variáveis sexo, série e idade, procedeu-se as respectivas análises. No que se refere às diferenças entre meninas e meninos, os resultados referentes aos instrumentos Bender-SPG, Reconhecimento de Palavras, Cloze e o ADAPE podem ser visualizados na Tabela 10.

Tabela 10 - Média desvio-padrão, *t* e *p* dos participantes divididos por gênero para todos os instrumentos

	Sexo	N	Média	DP	<i>t</i>	<i>p</i>
Bender Total	Masculino	153	7,73	3,90	1,647	0,101
	Feminino	144	7,02	3,43		
Reconhecimento de Palavras	Masculino	153	37,13	11,46	-2,580	0,010
	Feminino	144	40,43	10,51		
Cloze	Masculino	153	9,45	3,96	-1,472	0,142
	Feminino	144	10,12	3,83		
ADAPE	Masculino	153	35,42	25,99	3,218	0,001
	Feminino	144	26,58	20,90		

Com base nos dados acima dispostos, podemos verificar que houve diferença significativa entre os sexos no que se refere ao instrumento Reconhecimento de Palavras, sendo que as meninas tiveram melhor desempenho. Ao lado disso, no ADAPE, a diferença entre as médias de meninos e meninas também foi significativa, sendo que novamente as

meninas tiveram um desempenho superior ($M=26,58$) considerando-se que o ADAPE pontua os erros.

Os resultados estão em consonância com os encontrados no trabalho de Suehiro (2005), uma vez que os meninos obtiveram uma pontuação média maior no ADAPE (erro) do que as meninas, revelando uma diferença significativa entre eles. Em contrapartida, Santos (2006), em pesquisa de avaliação da maturidade percepto-motora e traços de personalidade, ao comparar os resultados em relação ao sexo no Bender-SPG, observou que não ocorreu diferença estatisticamente significativa ($t=1,424$; $p=0,156$), sendo a média dos meninos de 7,21 e das meninas 8,15.

Contrariando a superioridade feminina no que se refere ao aspecto gráfico, Machado (1978), ao investigar a influência das variáveis, idade, sexo e nível socio-educacional na capacidade de organização perceptivo-motora de crianças entre 7 e 10 anos, por meio do Teste de Bender verificou em seus resultados, que os meninos apresentaram uma discreta tendência a superioridade na reprodução gráfica. Embora essa diferença não tenha sido estatisticamente significativa.

Comparação de médias por série

Na presente pesquisa, os resultados apontaram um desempenho superior na média das crianças em relação às séries, em consonância com os estudos apontados até agora. Esses resultados podem ser visualizados na Tabela 11.

Tabela 11 - Comparação entre as médias de crianças de diferentes séries com base nos instrumentos por meio da ANOVA

	gl	<i>F</i>	<i>p</i>
Bender Total	3	2,484	0,061
Reconhecimento de Palavras	3	11,124	0,000
Cloze	3	6,298	0,000
ADAPE	3	4,057	0,008

Em relação às séries, tomaram-se como referência os grupos de 1^a a 4^a, os resultados revelaram que houve diferença entre eles em todos os instrumentos, embora no Bender-SPG, o nível de significância tenha sido marginal [$F=(3,297)= 2,484$; $p= 0,061$]. Ainda nesse sentido, o Reconhecimento de Palavras [$F=(3,297)= 11,124$; $p= 0,000$] e o Cloze [$F=(3,297)= 6,298$; $p= 0,000$] apresentaram resultados altamente significativos. . A Média dos instrumentos e a Série dos participantes, também podem ser visualizadas por meio das Figuras 1, 2 3 e 4.

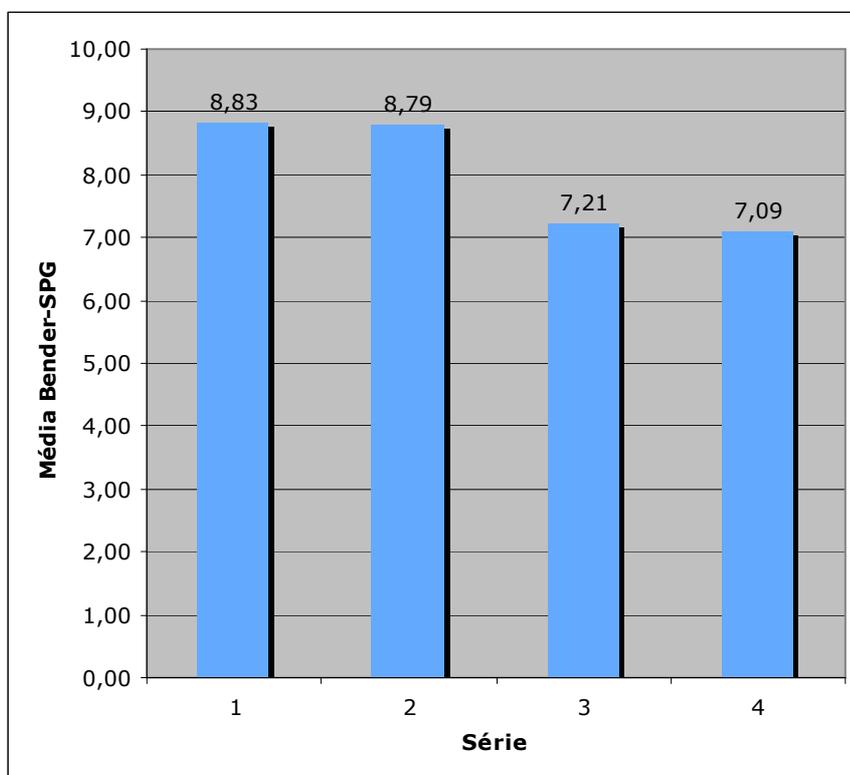


Figura 1 - Resultado da Média por Série no Teste Bender-SPG

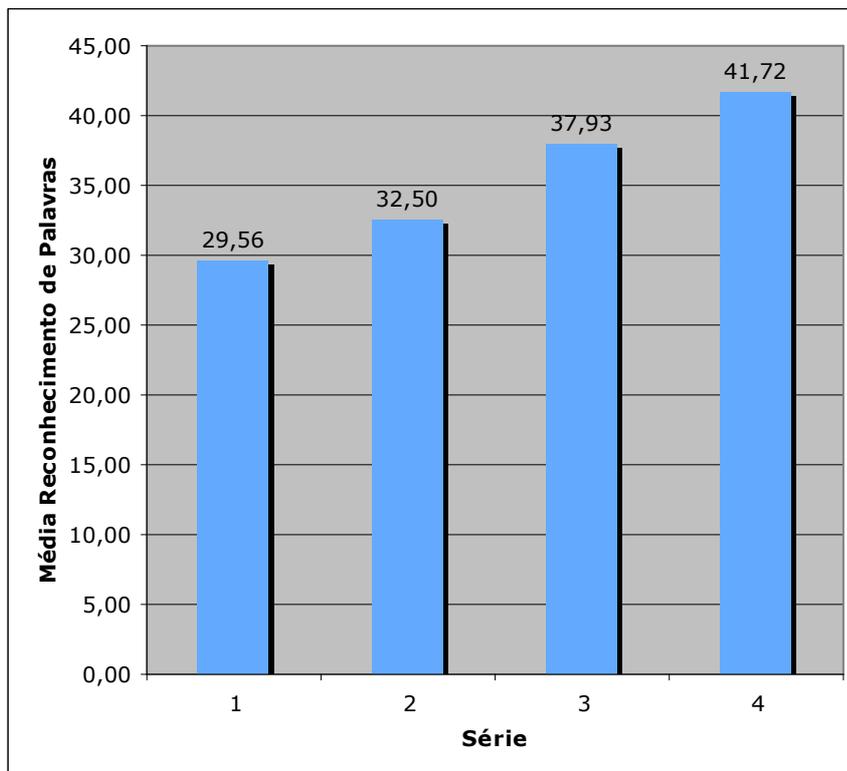


Figura 2 - Resultado da Média por Série no Teste Reconhecimento de Palavras

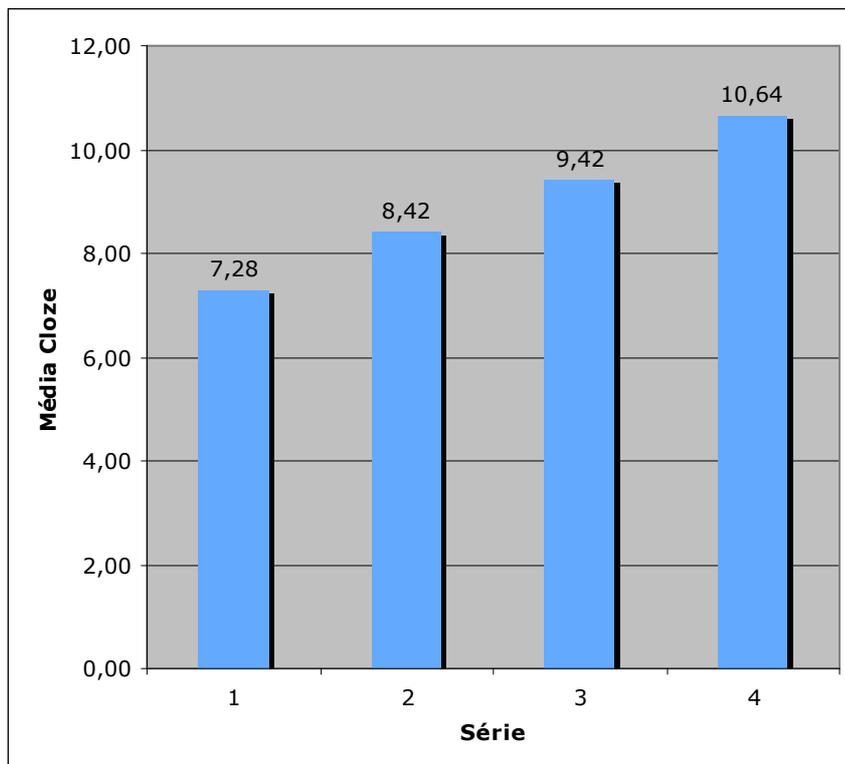


Figura 3 - Resultado da Média por Série no Teste em Compreensão de Leitura (Cloze)

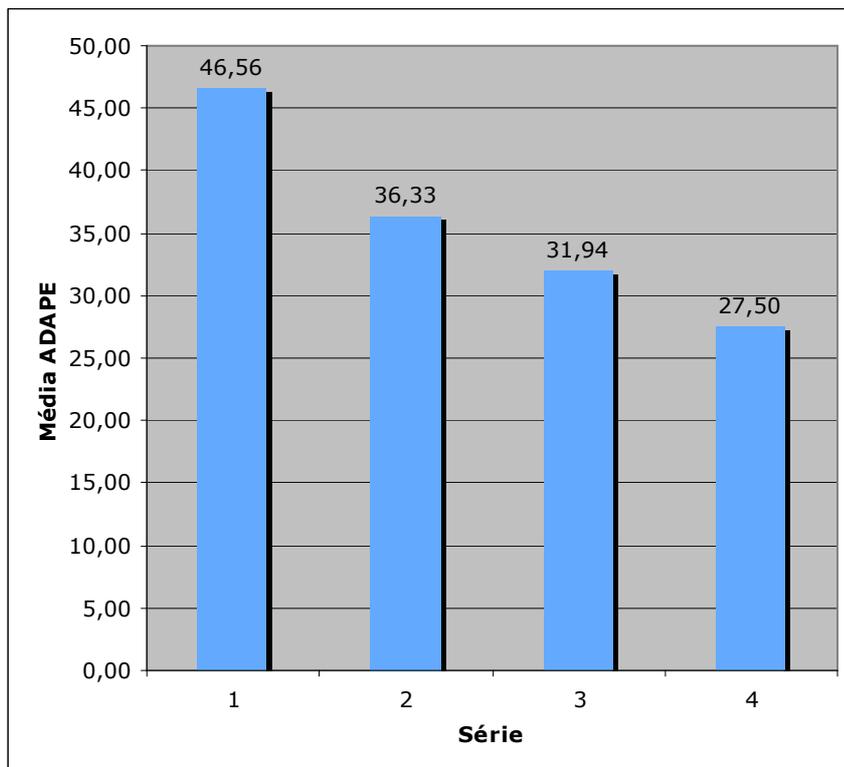


Figura 4 - Resultado da Média por Série no Teste ADAPE

Para verificar as médias dos resultados dos testes entre as variáveis séries, que confirmassem as diferenças encontradas, utilizou-se o teste Tukey (nível de significância de 0,05). Esses resultados estão ilustrados nas tabelas 12, 13, 14 e 15.

Tabela 12 - Subconjunto formado pela prova de *Tukey* em relação à pontuação total do Teste de Bender-SPG para a série

Série	N	Pontuação total Bender-SPG
		1
4ª série	137	7,09
3ª série	118	7,21
2ª série	24	8,79
1ª série	18	8,83

A comparação das médias entre as séries dos participantes da pesquisa revelou como esperado, que conforme os alunos são aprovados de uma série a outra, a pontuação no Bender-SPG, diminui, ou seja, os participantes de 1ª e 2ª séries apresentaram uma média de 8,8 erros enquanto que os de 3ª e 4ª séries apresentam uma média de 7,2 e 7,1 respectivamente. Esses dados confirmam a hipótese levantada por Koppitz (1989) sobre a importância do aspecto maturacional para um bom desempenho no Bender e evidentemente a relação idade/série. No entanto, houve a organização de um único conjunto, tal como apresentado na Tabela 12.

Tabela 13 - Subconjuntos formados pela prova de *Tukey* em relação à pontuação total do Teste Reconhecimento de Palavras para a série

Série	N	Reconhecimento de Palavras		
		1	2	3
1 ^a	18	29,56		
2 ^a	24	32,50	32,50	
3 ^a	118		37,93	37,93
4 ^a	137			41,72

Na Tabela 13, no que se refere à comparação de médias do Teste Reconhecimento de Palavras, o teste *Tukey* organizou o resultado em três grupos com dois elementos em comum em dois deles, sendo estes os alunos das segunda e terceira séries. Ocorreu uma clara diferenciação entre as médias da primeira e quarta séries.

Tabela 14 - Subconjuntos formados pela prova de *Tukey* em relação à pontuação total do Teste Cloze para a série

Escolaridade	N	Pontuação Total Cloze	
		1	2
1 ^a série	18	7,28	
2 ^a série	24	8,42	8,42
3 ^a série	118	9,42	9,42
4 ^a série	137		10,64

Como esperado, as médias dos alunos na pontuação total do Cloze aumentaram conforme a progressão na série, ou seja, os alunos da segunda série apresentaram um desempenho maior na média do que os alunos da primeira. Os da quarta série obtiveram média maior do que os da terceira. Os resultados foram organizados em dois grupos, com dois elementos comuns, são eles, os alunos da segunda e terceira séries. Esses dados demonstram que os grupos de 1ª e 4ª séries estão mais distantes em termos de conhecimentos da leitura e da escrita, o que parece pertinente, enquanto que os da 2ª e 3ª séries estão próximos e alfabetizados, supõe-se.

Tabela 15 - Subconjuntos formados pela prova de *Tukey* em relação à pontuação total do Teste ADAPE para a série

Escolaridade	N	Pontuação Total ADAPE	
		1	2
4ª série	137	27,50	
3ª série	118	31,94	
2ª série	24	36,33	36,33
1ª série	18		46,56

A Tabela 15 apresenta os subconjuntos formados pela prova de *Tukey* em relação à pontuação total do Teste ADAPE e a série. Houve uma diferenciação da primeira série em relação às demais. Isto significa que, quanto maior a série, a possibilidade de erros no ADAPE diminui porque nesse instrumento pontuam-se os erros, como já explicitado.

Comparação de médias por idade

Tabela 16 - Comparação entre as médias de crianças de diferentes idades com base nos instrumentos por meio da ANOVA

	gl	<i>F</i>	<i>p</i>
Bender Total	3	5,359	0,001
Reconhecimento de Palavras	3	8,069	0,000
Cloze	3	2,142	0,095
ADAPE	3	1,781	0,151

Em relação às idades, tomaram-se como referência os grupos de 7, 8, 9 e 10 anos. Os resultados revelaram que houve diferença entre eles nos instrumentos B-SPG [$F=(3,297)= 5,359; p= 0,001$] e Reconhecimento de Palavras [$F=(3,297)= 8,069; p= 0,000$]. A seguir, encontram-se as Figuras 5, 6, 7 e 8 com os resultados das médias do Teste de Bender-SPG, Reconhecimento de Palavras, Cloze, ADAPE e a Idade dos participantes.

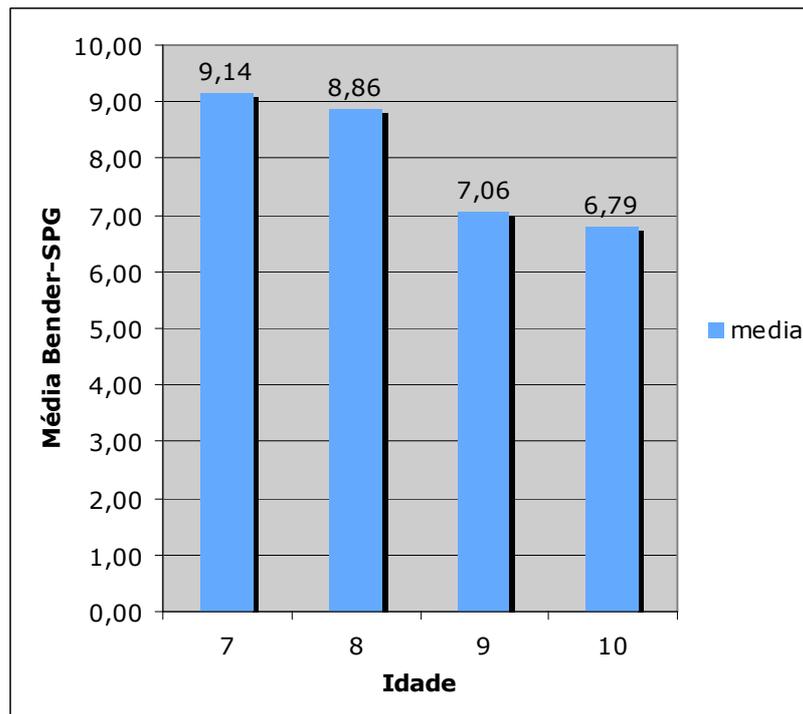


Figura 5 - Resultado da Média por Idade no Teste Bender-SPG

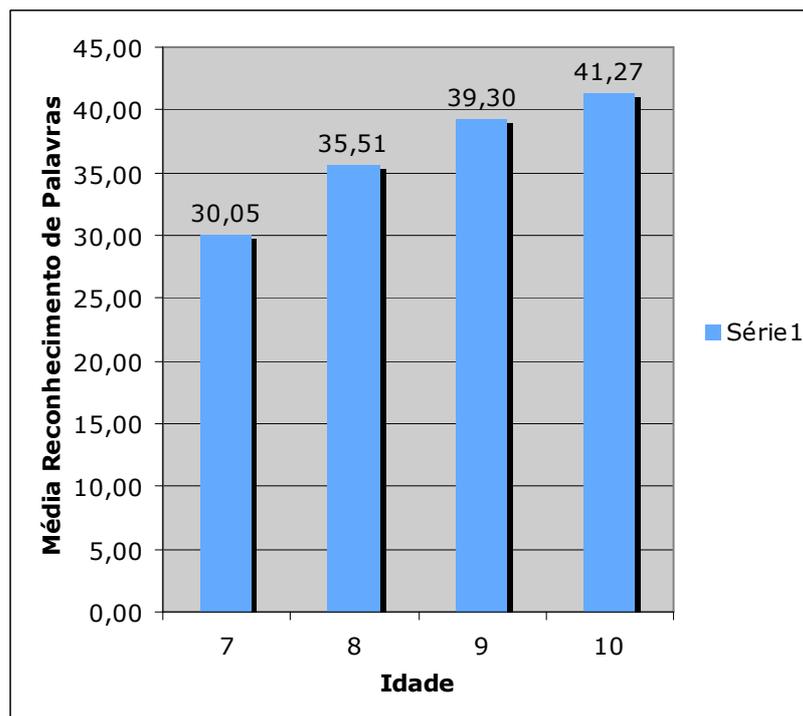


Figura 6 - Resultado da Média por Idade no Teste de Reconhecimento Palavras

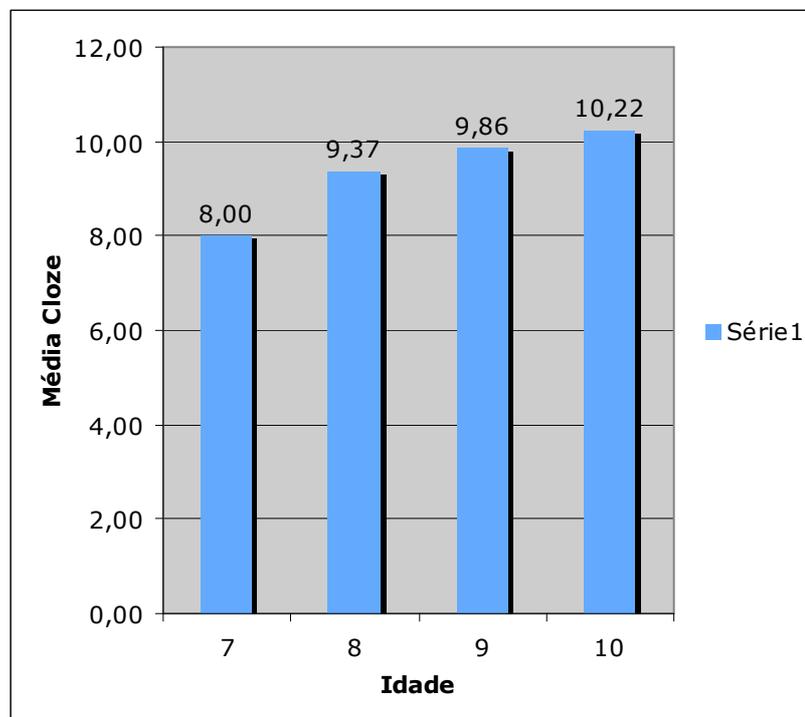


Figura7 - Resultado da Média por Idade no Teste de Compreensão em Leitura (Cloze)

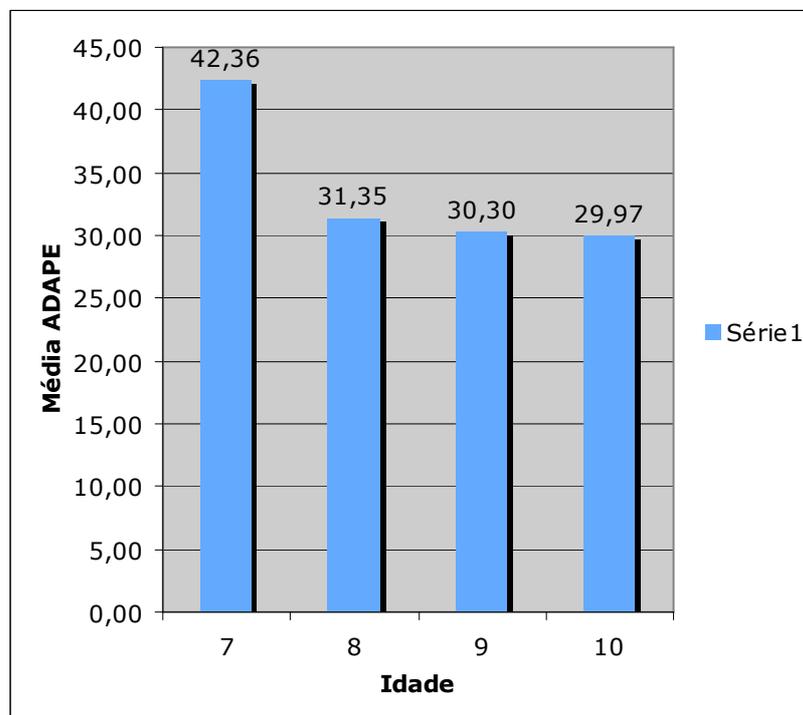


Figura 8 - Resultado da Média por Idade no Teste ADAPE

Para se verificar as médias dos resultados dos testes entre a variável idade, que justificassem as diferenças encontradas, utilizou-se o teste de *Tukey* (nível de significância de 0,05). Seus resultados são mais bem visualizados nas Tabelas 17, 18,19 e 20.

Tabela 17 - Subconjuntos formados pela prova de *Tukey* em relação à pontuação total do Teste de Bender para a idade

Idades	N	Pontuação total		
		Bender-SPG		
		1	2	3
10	100	6,79		
9	132	7,06	7,06	
8	43		8,86	8,86
7	22			9,14

O teste de *Tukey* separou as idades em relação à pontuação no Bender em três conjuntos e reforçou a idéia de que as crianças de 10 anos, portanto mais velhas, apresentam menos erros que as de 7 anos, confirmando a hipótese evolutiva do grafismo relacionado à maturidade percepto-motora, ou seja, evidências de validade desenvolvimental (Machado, 1978; Mazzeschi & Lis, 1999; entre outros).

Tabela 18 - Subconjuntos formados pela prova de *Tukey* em relação à pontuação do Teste Reconhecimento de Palavras para a idade

Idade	N	Pontuação total Reconhecimento de Palavras	
		1	2
7	22	30,05	
8	43	35,51	35,51
9	132		39,30
10	100		41,27

O teste de *Tukey* separou as idades em relação à pontuação no Reconhecimento de Palavras em dois grupos, revelando que não houve evidências de igualdade entre as médias dos resultados no teste e a variável idade. Pode-se verificar que, à medida que a idade avançava, a média dos acertos aumentava. Esses resultados também estão em consonância com as pesquisas apontadas neste estudo (Bandeira & Huttz, 1994; Oliveira, Sisto, Souza, Brenelli & Fini, 1994).

Tabela 19 - Subconjuntos formados pela prova de *Tukey* em relação à pontuação do Teste de Compreensão em leitura (Cloze) para a idade

Idade	N	Pontuação Total Cloze	
		1	2
7	22	8,00	
8	43	9,37	9,37
9	132	9,86	9,86
10	100		10,22

Quanto à compreensão de leitura, o teste de *Tukey* organizou as idades em dois grupos. Os resultados revelaram que, à medida que a idade avança, melhora o desempenho da criança no teste, fato que era esperado, considerando que a criança, ao progredir da 1ª para a 4ª série, normalmente apresenta melhora no desempenho da leitura.

Tabela 20 - Subconjuntos formados pela prova de *Tukey* em relação à pontuação do Teste ADAPE para a idade

Idade	N	Pontuação total ADAPE
		1
10	100	29,97
9	132	30,07
8	43	31,35
7	22	42,36

No que se refere aos instrumentos Cloze e ADAPE, como pode ser observado nas Tabelas 17 e 18, respectivamente, as médias dos instrumentos apresentaram diferença quando a idade é levada em consideração. Para o teste de Cloze, a média cresce quando a idade aumenta. No ADAPE, a média decresce, considerando que se pontuam os erros nesse instrumento. Esses resultados vêm ao encontro dos estudos de Shapiro e Simpson (1995), que concluíram que a idade estava relacionada ao desenvolvimento. Uma indicação de que

as habilidades visuais e motoras continuam evoluindo para além da idade de 11 anos, uma vez que, em sua pesquisa, os autores avaliaram adolescentes.

Comparação de médias por tipos de escola

As crianças que estudam em escolas particulares apresentaram um desempenho melhor na média dos resultados dos instrumentos aplicados quando comparadas com a média dos resultados das crianças que freqüentam as escolas públicas. Esses resultados podem ser verificados nas Figuras apresentadas a seguir.

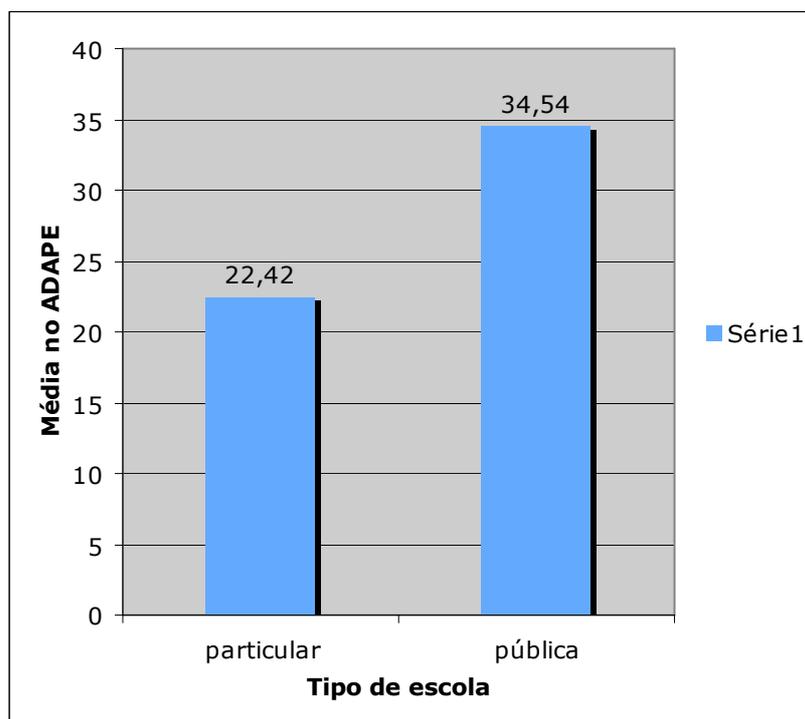


Figura 9 - Resultado da Média por escola no Teste ADAPE

Na comparação entre a média dos alunos de diferentes escolas com o Teste ADAPE, pode-se notar que a média dos alunos das escolas particulares foi de 22,42, portanto com menos erros do que a média dos alunos de escolas públicas (34,54) apontando um melhor desempenho dos alunos das escolas particulares. O mesmo não ocorreu entre a média dos alunos dessas escolas com o Teste de Bender, que não apresentou diferença estatisticamente significativa na média dos alunos em relação às escolas, demonstrando, mais uma vez, que o resultado dele decorrente está mais diretamente relacionado com a maturidade percepto-motora do indivíduo.

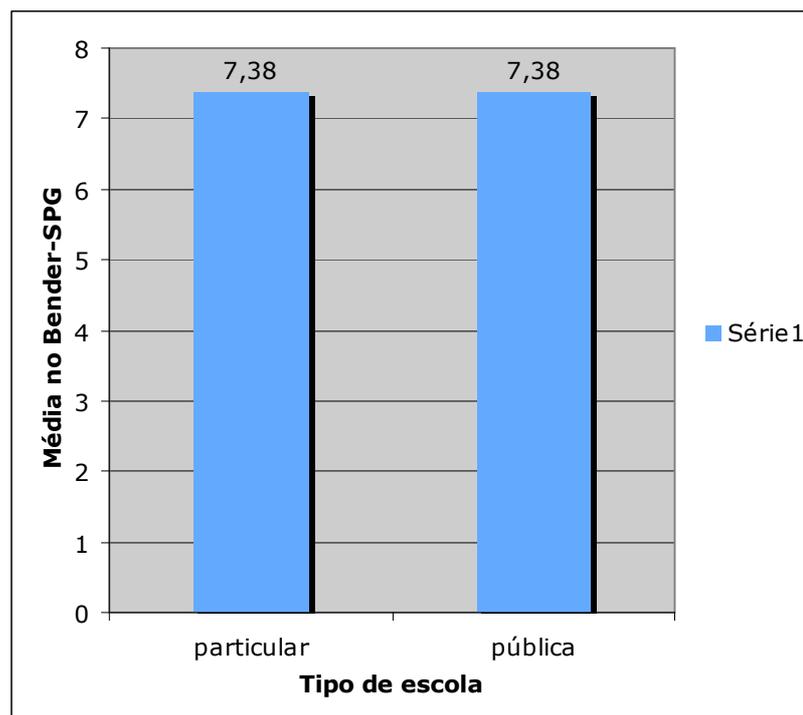


Figura 10 - Resultado da Média por escola no Teste de Bender-SPG

Na comparação entre as médias dos alunos das escolas particulares e escolas públicas, no que se refere ao Teste de Cloze, os resultados revelaram-se estatisticamente significativos: 11,73 entre os alunos das escolas particulares e 9,04 entre os das escolas públicas. O desempenho de alunos das escolas particulares foi melhor, conforme demonstra a Figura 11.

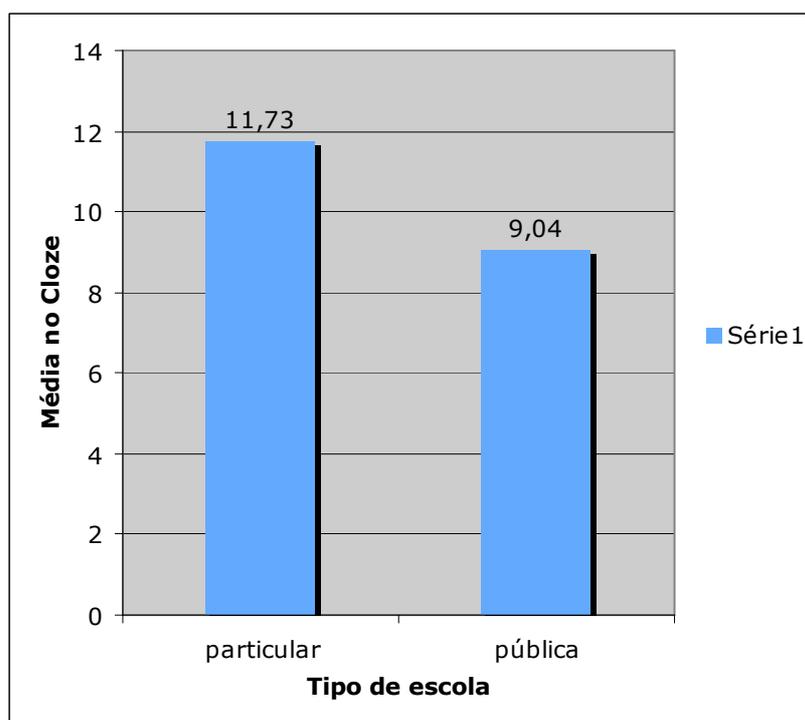


Figura 11 - Resultado da Média por escola no Teste de Compreensão em leitura (Cloze)

A Figura 12 mostra que a média dos alunos de escolas particulares também seguiu a tendência da média da maioria dos instrumentos *versus* escolas, ou seja, as crianças das escolas particulares apresentaram um desempenho significativamente maior na média (43,9) quando comparados com os alunos das escolas públicas (36,79).

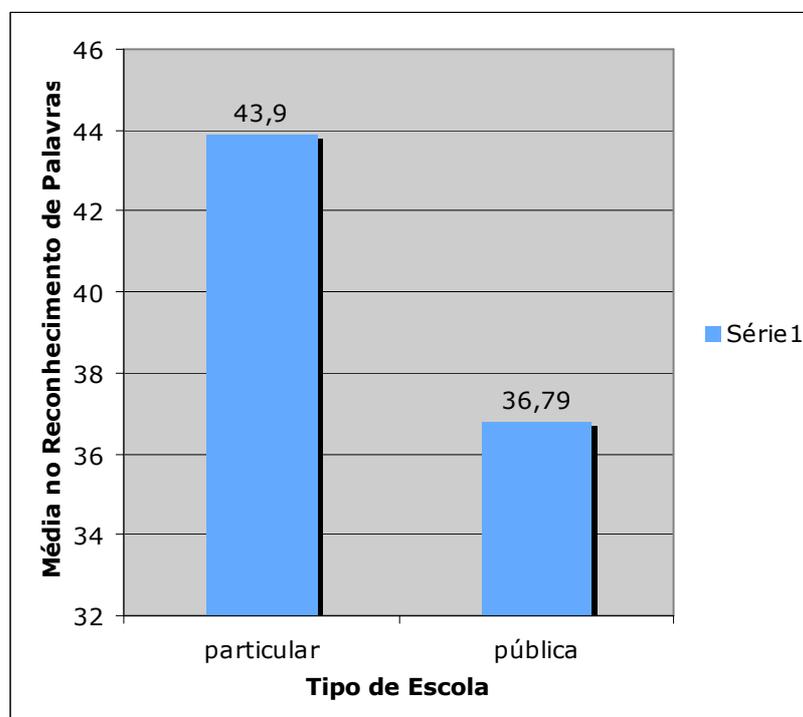


Figura 12 - Resultado da Média por escola no Teste Reconhecimento de Palavras

Estes resultados também encontram respaldo na pesquisa de Machado (1978), que, para atender as finalidades do estudo, distribuiu os participantes em duas amostras, sendo

que a primeira foi composta por 397 crianças e se destinou à verificação da influência da idade e do sexo nos resultados. Já a segunda, com 132 crianças, buscou verificar até que ponto o nível socio-educacional (NSE), estabelecido com base no nível ocupacional e educacional dos pais, interferiu na realização gráfica. Os resultados demonstraram que o grupo mais carente em termos sócio-educacionais revelou nítido atraso no domínio das noções espaciais e desempenho inferior quanto à evolução da organização perceptivo-motora.

Isso está em consonância com Kroeff (1992), que, ao apresentar o resultado da aplicação do Teste de Bender em mais de 1000 crianças de escolas públicas e particulares, entre 5 e 10 anos e 11 meses, destacou a importância do uso de normas de pontuação do teste diferenciadas em razão do tipo de escola freqüentada. Ao analisar a influência de diferentes condições socioeconômico-culturais sobre o desempenho de crianças, do Jardim a 5ª série, verificou que tanto em relação a variável idade, quanto ao nível de escolaridade, as crianças de escola particular apresentaram um desempenho superior ao das que freqüentavam a escola pública. Mesmo sendo, em média, mais velhas que as de escolas particulares, apresentavam desempenho inferior ao das mais novas de escolas particulares que freqüentavam a mesma série.

O desempenho mais baixo das crianças de nível sócio-econômico inferior também foi evidenciado pelo estudo realizado por Pinelli Jr. e Frey (1991). Os pesquisadores aplicaram as Matrizes Progressivas Coloridas de Raven, o Teste Gestáltico de Bender e o *Beery Developmental Test of Visual-Motor Integration* em 415 crianças brasileiras, das quais 46,3% eram do sexo feminino e 53,7% do masculino, sendo 206 de uma escola

pública da periferia de Brasília e 209 de uma escola particular da área urbana, correspondendo a dois níveis sócio-econômicos. As crianças avaliadas tinham entre 7 e 12 anos e foram divididas em quatro faixas etárias. Os resultados indicaram que os escores obtidos pelas crianças de nível sócio-econômico mais baixo (escola pública) foram menores que os das crianças de níveis economicamente superior (escola particular) em todas as medidas e que as diferenças entre os dois grupos foram pequenas, embora significativas, para o *Beery Test*.

A classe social e a série escolar das crianças influenciaram os escores nos três testes mais do que a idade, sendo que o Raven foi o instrumento mais influenciado pelas diferenças de classe social. Evidenciou-se também uma correlação moderada (0,47 a 0,69) entre as habilidades de integração visomotora e de discriminação visual, especialmente entre os estudantes da escola particular. Os autores enfatizaram que as diferenças maturacionais em função da idade e da série escolar foram melhor estimadas pelo *Beery* do que pelo Raven ou pelo Teste Bender, e que a variação das faixas etárias para cada série e o número menor de protocolos das séries escolares superiores dificultaram a interpretação dos dados obtidos (Pinelli Jr. & Frey 1991).

Feita essas considerações e com base nessa nova proposta de correção, tendo em vista que os resultados deste estudo confirmam que a maturidade percepto-motora está relacionada com a compreensão de leitura e com a escrita verificou-se que há evidências de validade de construto convergente-discriminante (Anastasi & Urbina, 2000). O mesmo desempenho ocorreu em relação às séries, idade, e escolas estudadas, o que lhe confere

evidências de validade de critério e possibilita uma avaliação da população infantil mais condizente com a realidade cultural brasileira.

Os resultados referentes às diferenças de médias entre as variáveis sexo, série, idade e escolas estão em consonância com pesquisas já realizadas com o instrumento Bender-SPG, e citadas nesta pesquisa (Suehiro, 2005; Santos, 2006; Bartholomeu, 2006, dentre outras). Um outro dado apontado pelo presente estudo está relacionado à relação entre os construtos e as variáveis idade e escolaridade. Embora os coeficientes tenham se revelados baixos, foram significativos. Dentre eles podemos citar o da escolaridade e o Teste de Reconhecimento de Palavras, que foi a maior correlação encontrada.

Quanto à busca de evidências de validade de critério, podemos citar também a validade de critério para o Bender-SPG pela comparação de grupos extremos, ou seja, pela comparação de grupos divididos por melhores e piores resultados nos demais instrumentos utilizados na pesquisa. Esses resultados demonstraram mais uma vez, os pressupostos de que o desenvolvimento psicomotor é condição essencial para as habilidades de leitura e escrita, já citada anteriormente, e reforçada por pesquisas apontadas neste estudo (Sisto e cols, 2005).

Para Limomgi (1996), o processo de aprendizagem está baseado em três princípios neurológicos fundamentais, a saber, a atividade motora, a praxia (movimentos ordenados em seqüência e organizados para exercer uma determinada finalidade) e o sistema labiríntico-tônico-postural que permite a interação dos outros dois princípios de maneira adequada. Outro fator de relevante importância são os mecanismos envolvidos na realização das tarefas acadêmicas das crianças na fase escolar ou da aprendizagem da

leitura e da escrita, tais como, orientação espaço-temporal, lateralidade estruturada, domínio adequado do gesto e do instrumento e dominância bem definida. Neste sentido, (Le Boulch, 1984; Moraes, 1986; e Oliveira, 2001) demonstraram em seus estudos a influência do aspecto psicomotor na aprendizagem escolar.

Nessa mesma direção, Moraes (1986) aponta para o fato de que as crianças devam reunir as condições para a alfabetização (processo de leitura e escrita) no período pré-escolar, e que estas, nessa fase, necessitam adquirir as habilidades básicas necessárias ao período alfabetizante. Essas habilidades envolvem desde o conhecimento adequado do seu corpo (abrangendo o esquema e a imagem corporal) até o domínio da lateralidade (uso preferencial de um lado do corpo) e noções de direita e esquerda.

Esses indicativos foram demonstrados por Oliveira (1992) em seu estudo com 20 participantes de uma escola pública, que apresentavam dificuldades de aprendizagem. Os resultados evidenciaram uma melhora significativa no desempenho em leitura e escrita das crianças que foram submetidas ao processo de intervenção com exercícios psicomotores, o que levou a autora à hipótese de que um bom desempenho psicomotor contribui para a melhora do desempenho escolar.

Nesse mesmo sentido, os estudos de Sisto (1994) e Fonseca (1995) revelaram a importância dos aspectos motores para a aprendizagem da escrita e consideraram esses fatores relacionados às dificuldades de aprendizagem e ao desempenho escolar. A habilidade percepto-motora tem sido destacada como condição importante para a aprendizagem, considerando que tais estudos enfatizam a psicomotricidade como um aspecto significativo para o aprendizado e que um problema psicomotor pode gerar

dificuldades na discriminação de letras, na atividade de coordenação motora fina, dentre outros.

A partir desses pressupostos, verificou-se que o desenvolvimento perceptivo-motor adequado é requisito para o desenvolvimento de habilidades e, especialmente, a de leitura e escrita. Um dos instrumentos que tem sido considerado útil para o diagnóstico diferencial de crianças com problemas de aprendizagem é o Teste Gestáltico de Bender, mas em sua nova versão, Bender-SPG de Sisto, Noronha e Santos (2005) que, como apontou o presente estudo, possui os requisitos psicométricos necessários para a avaliação de tais crianças.

Nesse sentido, tanto em estudos estrangeiros, como entre os estudos brasileiros, recuperados nesta pesquisa, houve um grande número deles que apontaram para a necessidade de se buscar evidências de validade para muitos dos usos do Bender e, especialmente, do sistema de avaliação desenvolvido por Koppitz (Taylor, Kauffman & Partenio, 1984; Weinstein, 1993; Shapiro & Simpson, 1995; Moura, 1982; Puente & Maciel Jr. 1984; Pinelli Jr., & Pasquali, 1990; Sisto, Santos & Noronha, 2005 dentre outros).

O desenvolvimento das crianças se dá a partir de suas experiências sensório-motoras nos primeiros anos de vida e, conforme Bender (1938), a maturidade percepto motora é um requisito para o desenvolvimento das habilidades acadêmicas, dentre elas a leitura e a escrita. Neste contexto, a avaliação das dificuldades de aprendizagem se constitui como uma necessidade crescente. É preciso considerar que muitos são os fatores que influenciam as crianças durante o processo de construção da escrita.

Os questionamentos com o Teste de Bender tornaram-se motivos de investigação, desde a sua forma de aplicação, população atendida, até, e, principalmente o sistema de correção.

Este último fomentou muitas pesquisas, tanto no que se refere ao quesito “teste de maior conhecimento e uso” (Koppitz 1975), quanto à preocupação na elaboração de outros sistemas como os de Billingslea (1948), Hutt (1969), Pascal e Suttell (1951). Neste contexto, esta pesquisa buscou verificar se o Bender-Sistema de Pontuação Gradual (B-SPG) pode superar as fraquezas identificadas como sensíveis para a avaliação do Bender.

CAPÍTULO V

CONSIDERAÇÕES FINAIS

Esse estudo pretendeu buscar evidências de validade convergente-discriminante, por meio da correlação do Teste de Bender-SPG com os instrumentos Escala de Avaliação de Dificuldade na Aprendizagem da Escrita (ADAPE), Teste de compreensão em leitura (Cloze) e o Teste de Reconhecimento de Palavras. Buscou também evidências de validade de critério para o Bender-SPG pela comparação de grupos divididos por melhores e piores resultados nos instrumentos Cloze, ADAPE e Reconhecimento de Palavras. Ainda verificou a Precisão do B-SPG, além de se analisar as diferenças de médias dos instrumentos entre os participantes no que se refere às variáveis idade, sexo, série e tipos de escola.

Percebeu-se que o Teste de Bender pode ser utilizado como um instrumento de avaliação, não somente para verificar a maturidade percepto-motora, mas também como um recurso que acrescenta a um amplo processo de avaliação. Os estudos com o novo sistema de correção foram necessários, pois embora ele esteja em consonância com os requisitos ressaltados pelo Conselho Federal de Psicologia, no que se refere ao uso e comercialização, pesquisas com enfoques e amostras distintas tendem a aprimorar o instrumento.

Como se pode verificar neste trabalho, muitos foram os pesquisadores que se dedicaram ao estudo do Bender nas suas mais variadas formas de correção. Em acréscimo, parece não ser novidade o fato de que ele foi um dos instrumentos mais utilizados tanto no

estrangeiro como no Brasil. Os estudos apresentados foram variados; alguns apontando o seu uso para a área de dificuldade de aprendizagem, outros envolvendo dados estatísticos para verificar a diferença no desempenho entre populações masculinas e femininas e alguns se ocupando da investigação de pacientes lesionados, dentre outros (Lesiak, 1994; Machado, 1978; Suehiro, 2005).

No que se refere aos objetivos da pesquisa, evidências de validade e a estimativa da precisão puderam constatados. Em relação à primeira, evidências de validade, a pesquisa apontou que o Bender-SPG apresentou evidências de validade de construto convergente-discriminante e evidências de validade de critério. Quanto ao índice de precisão do instrumento, revelou-se significativo, resultado que corrobora com a necessidade de se construir instrumentos fidedignos e que considerem os pressupostos necessários na utilização de testes psicológicos, como prescreve o Conselho Federal de Psicologia (CFP) por meio da Resolução 002/2003. Dessa forma, o compromisso do psicólogo torna-se maior ao avaliar os sujeitos com queixa de dificuldades de aprendizagem, bem como ao propor alternativas viáveis para que os alunos possam superar eventuais dificuldades.

É importante salientar também, que em função das controvérsias relacionadas à aprendizagem infantil no setor educacional, faz-se necessária a presença de um profissional habilitado, com boa formação e experiência na área de avaliação psicológica. Além disso, o psicólogo precisa estar atento ao escolher os instrumentos que serão incluídos no processo de avaliação infantil, verificando a validade e a precisão do teste, e se este apresenta facilidade na aplicação, bem como, se favorece a compreensão e a execução das tarefas para os testandos.

Observou-se, que tais modalidades foram contempladas no referido teste. Uma outra característica do teste está relacionada ao aspecto motivacional, principalmente, quando se trata de instrumentos utilizados para a avaliação infantil. O caráter lúdico, a rapidez na aplicação e as opções de uso de forma individual ou coletiva são outros fatores positivos na utilização do B-SPG. Essa vertente inovada na modalidade de aplicação torna o instrumento com possibilidades de menor exposição da criança frente a situações de avaliação que, em muitos casos, geram constrangimento e desmotivação.

Vale destacar, no entanto, que embora o Bender-Sistema de Pontuação Gradual (B-SPG), de Sisto, Noronha e Santos (2005), seja uma proposta inovadora e promissora de avaliação, outras evidências de validade devem ser pesquisadas. Como por exemplo, aplicando-o em amostras de outros estados brasileiros, bem como o correlacionando com outros instrumentos de construtos relacionáveis. A realização de novos estudos com a utilização desse sistema poderá propiciar a ampliação do conhecimento das características psicométricas do instrumento e complementar as lacunas e sanar as limitações deste trabalho.

É sabido que, por mais esforços que o pesquisador empregue no desenvolvimento de uma pesquisa, é inevitável a ocorrência de problemas no decorrer da mesma. Nem sempre é possível controlar todas as variáveis desejadas. Vale lembrar que no procedimento de coleta de dados, em função das dificuldades de aceitabilidade dos pais, principalmente das escolas particulares, em assinar a autorização necessária para o início dos trabalhos, levou a uma falta de homogeneidade entre o número de protocolos, idade e as séries focalizadas.

Em relação aos participantes das 1ª série (N=18) tomou-se o cuidado de coletar os dados no final do ano letivo, no sentido de evitar a falta de contato com a leitura e escrita. Vale destacar que os testes de Compreensão em Leitura (Cloze) e o ADAPE requerem conhecimento e familiaridade em codificar e decodificar os símbolos da língua portuguesa. Espera-se ainda que, outros estudos, além dos citados, referentes à investigação do Bender, relacionando-o a outras variáveis como problemas neuropsicológicos em crianças, validade do mesmo em pacientes psiquiátricos ou estudos relacionados a outras faixas etárias (Blanchard, 1998; Maciel Jr. & La Puente, 1993), possam também ser desenvolvidos com o sistema Bender-SPG.

V I - REFERÊNCIAS

- Adánez, G. A. (1999). Procedimientos de construccions y análisis de tests psicométricos. Em S. M. Wechsler & R. S. L. Guzzo, (Orgs.), *Avaliação Psicológica: Perspectiva Internacional*. (pp.57-100). São Paulo: Casa do Psicólogo.
- American Educational Research Association, the American Psychological Association & National Council on Measurement in Education, (1999). *The Standards for Educational and Psychological Testing*. Washington: American Educational Research Association.
- Anastasi, A. (1967,1977a) *Testes Psicológicos* – EPU, São Paulo.
- Anastasi, A., & Urbina, S. (2000). *Testagem Psicológica*. Porto Alegre: Artes Médicas.
- Aylward, E. H., & Schmidt, M. A. S. (1986). An examination of the three tests of visual motor integration. *Journal of Learning Disabilities*, 19 (6), 328-330.
- Bandeira, R. D., & Hutz, C. S. (1994). A contribuição dos testes DFH, Bender e Raven na predição do rendimento escolar na primeira série. *Psicologia: Teoria e Pesquisa*, 10 (1), 59-72.
- Bartholomeu, D. (2006). *Teste Gestáltico Visomotor de Bender e Desenho da Figura Humana: Convergências de Avaliação?* Dissertação de Mestrado da Universidade São Francisco, Itatiba, São Paulo.
- Bazi, G. A. P., & Sisto, F. F. (2006). Alegria, tristeza medo e coragem em crianças com dificuldades de aprendizagem. Em F. F. Sisto & S. C. Martinelli (Org.), *Afetividade e Dificuldades de Aprendizagem*. (pp.57-72). São Paulo: Vetor.

- Bender, L. (1974). *Test Gestáltico Visomotor: Usos y Aplicaciones Clínicas*: (Tradução D. Carnelli). Buenos Aires: Editorial Paidós.
- Bender, L. (1969). *Test Gestáltico Visomotor* de Bender 1a edição, (Trad. D. Carnelli) Bibl. da Psicometria e Psicodiagnóstico, Série Mayor, vol.7, Buenos Aires, Editorial Paidós.
- Billingslea, F.Y. (1948). The Bender-Gestalt: an Objective Scoring Method and Validating Data. *Journal of Clinical Psychology*, 41-27.
- Bispo, N. L. (2000). *Imagem Mental, memória e dificuldades de aprendizagem na escrita*. Dissertação de Mestrado, UNICAMP, Campinas.
- Bormuth, J. R. (1968). Cloze Test Readability: Criteria Reference Score. *Journal Educational Measurement*, 5, 189-196.
- Bucklev, P.D. (1978). The Bender Gestalt Test: A review of reported research with school-age subjects, 1966-1977. *Psychology in the School*, 15 (3), 327-335.
- Bunchaft, G. & Cavas C.S.T. (2002). *Sob Medida: um guia sobre a elaboração de medidas de comportamento e suas aplicações*. São Paulo: Vetor.
- Blanchard, T. F. (1998). An abbreviated Bender Gestalt test for detection of neurological dysfunction. *Dissertation Abstracts International*, 58(8), 4437.
- Brannigan, G. G., & Brunner, N. A. (1989). *The modified version of the Bender-Gestalt Test for preschool and primary school children*. Brandon-VT: Clinical Psychology Publishing Co.
- Brannigan, G. G., & Brunner, N. A. (1991). Relationship between two scoring systems for the modified version of the Bender Gestalt Test. *Perceptual and Motor Skills*, 72, 286.

- Brannigan, G. G., & Brunner, N. A. (1993). Comparison of the Qualitative and Developmental Scoring Systems for the Modified Version of the Bender-Gestalt Test. *Journal of School Psychology, 31*, 327-330.
- Brannigan, G. G.; Aabye, S. M.; Baker, L. A., & Ryan, G. T. (1995). Further validation of the qualitative scoring system for the modified Bender-Gestalt Test. *Psychology in the Schools, 32*(1), 24-26.
- Brannigan, G. G., & Brunner, N. A. (2002). *Guide to the Qualitative Scoring Systems for the Modified Version of the Bender-Gestalt Test*. Springfield, US: Charles C. Thomas, Publisher.
- Britto, G. N. O., & Santos, T. R. (1996). The Bender Gestalt Test for 5-to 15-year old Brazilian Children: norms and validity. *Jornal de Medicina e Biologia, 29* (11), 1513-1518.
- Brown, G. T.; Rodger., & Davis, A. (2003). Test of visual perceptual skills – Revised: An overview and critique: *Journal of OocupationTherapy, 10*:3-15.
- Canadian Psychological Association (1996). Guidelines for Education and Psychological Testing. Ontário, CA: CPA. Disponível On Line em: [http:// www.cpa.ca/guide9.html](http://www.cpa.ca/guide9.html) . Acessado 25/10/ 2005.
- Carneiro, G. R. S., Martinelli, S. C., & Sisto, F. F. (2003). Autoconceito e dificuldades de aprendizagem na escrita. *Psicologia: Reflexão e Crítica, 16*(3), 427-434.
- Conselho Federal de Psicologia CFP (2003). *Resolução Nº 002/2003*. Disponível em <http://www.pol.org.br> . Acessado em 12/09/2005.

- Conselho Federal de Psicologia CFP (2001). *Resolução No 025/2001*. Disponível em <http://www.pol.org.br>. Acessado em 14/09/2005.
- Cunha, J. A. (2000). *Psicodiagnóstico – V*. Porto Alegre: Artes Médicas.
- Chan, P. W. (2001). Comparison of visual motor development in Hong Kong and the USA assessed on the Qualitative Scoring System for the Modified Bender-Gestalt Test. *Psychological Reports, 88*, 236-240.
- Clawson, A. (1999). *Manual de Diagnóstico Clínico*. Porto Alegre: Artes Médicas, 7ª. Edição.
- Cronbach, L. J. (1996). *Fundamentos da Testagem Psicológica*. Porto Alegre: Artes Médicas.
- Field, K., Bolton, B., & Dana, R. H. (1982). An evaluation of three Bender-Gestalt scoring systems as indicators of psychopathology. *Journal of Clinical Psychology, 38*, 838-842.
- Finch, Jr. A. J., Spirito, A., Garrison, S., & Marshall, P. (1983). Developmental differences in Bender-Gestalt recall of children with learning and behavior problems. *Perceptual and Motor Skills, 56*, 87-90.
- Freire, T., & Almeida, L.S. (2001). Escalas de Avaliação: construção e validação. Em E. M. Fernandes, & L.S. Almeida (orgs). *Métodos e Técnicas de Avaliação*. Minho, Portugal: Universidade do Minho, Centro de Estudos em Educação e Psicologia.
- Gomes, M. A. M., & Boruchovitch, E. (2005). Desempenho no Jogo, Estratégias de Aprendizagem e Compreensão na Leitura. *Psicologia: Teoria e Pesquisa, 21*(3), 319-326.

- Hutt, M. L., & Briskin, G. J. (1960). *The Hutt Adaptation of the Bender Gestalt Test*. New York: Grune & Stratton Inc.
- Hutt, M. L. (1969). *The Hutt adaptation of the Bender Gestalt Test*. New York: Grune & Stratton.
- International Test Commission (2000). Guidelines on Adapting Tests. International Test Commission. Disponível On-line em [http:// www.Intestcom.org](http://www.Intestcom.org). Acessado em 26/08/2006.
- Kingeski, M.F. (2002). *Avaliação do Reconhecimento de palavras em crianças*. Dissertação de Mestrado, Programa de Estudos Pos Graduados em Psicologia. USF, Bragança Paulista, SP.
- Kingeski, M. F., & Sisto, F. F. (2004). Reconhecimento de Palavras nas séries iniciais do ensino fundamental. *Teoria e Prática da Educação*. 7 (2), 183-190.
- Koppitz, E.M. (1963). *The Bender Gestalt Test for young children*. New York: Grune, & Stratton.
- Koppitz, E. M. (1968). *Psychological evaluation of children's human figure drawings*. New York: Grune & Stratton.
- Koppitz, E. (1975). *The Bender Gestalt Test for young children: Research and application 1963-1973*. New York: Grune, & Stratton.
- Koppitz, E. M. (1989). *O Teste Gestáltico de Bender para crianças*. Porto Alegre: Artes Médicas.
- Kline, P. (1993). *The Handbook of Psychological Testing*. London: Rout ledge.

- Kroeff, P. (1988). Normas brasileiras para o Teste de Bender. *Psicologia: Reflexão e Crítica*, 1/2 (3), 12-19.
- Kroeff, P. (1992). Desempenho de crianças no Teste de Bender e nível sócio-econômico-cultural. *Psicologia: Reflexão e Crítica*, 5 (2), 119-126.
- La Puente, M., & Maciel, Jr., J. (1984). Procedimentos Operacionais na avaliação do Teste de Bender Infantil. *Estudos de Psicologia*, (3/4), 76-92.
- Lacks, P. (2000). Visuoconstructive abilities. In G. Groth-Marnat (Org.), *Neuropsychological assessment in clinical practice: A guide to test interpretation and integration*. (pp.401-436). New York: John Wiley & Sons.
- Lacks, P., & Newport K. (1980). A comparison of scoring systems and level of score experience on the Bender Gestalt Test. *Journal of Personality Assessment*, 44, 351-357.
- Ladesma, R., Adánez, G. M., & Mora, P.V. (2002). Análisis de consistência interna mediante Alfa de Cronbach: um programa baseado em gráficos dinâmicos. *Psico-USF*, 7(2), 143-152.
- Lee, S. Y., & Oh, S. W. (1998). Visuoperceptual and constructive ability disturbances of patients with traumatic brain injury in Hutt Adaptation of the Bender Gestalt Test. *Korean-Journal- of- Clinical-Psychology*, 17(1), 311-317.
- Lesiak, J. (1984), The Bender Visual Motor Gestalt test: Implications for the diagnosis and prediction of reading achievement. *Journal of School Psychology*, V. 22 pp. 391-405.
- Limongi, S. C. O. (1996). Da ação à comunicação: um processo de aprendizagem. *Revista Psicopedagogia*, 15 (36), 24-28.

- Van Kolck, O. L. (1977). *Técnicas de exame psicológico e suas aplicações no Brasil*. Petrópolis: Vozes,
- Lowndale, W. S., Rogers, B.J., & McCall, J. N. (1989). Concurrent validation of Hutt's Bender Gestalt screening method for schizophrenia, depression, and brain damage. *Journal of Personality Assessment*, 53 (4), 832-836.
- Machado, M.C. L. (1978). *Uso do Teste de Bender para avaliar a organização perceptiva motora de escolares paulistas*. Dissertação de Mestrado, Pontifícia Universidade Católica, São Paulo.
- Manarini, A.S. (1996). *Influência do sistema de escrita do português nos processamentos perilexicais de leitura*. Dissertação de Mestrado, Instituto de Psicologia, Universidade de São Paulo, SP.
- Mazzeschi, C., & Lis, A. (1999). The Bender-Gestalt test: Koppitz's Developmental Scoring System administered to two samples of Italian preschool and primary school children. *Perceptual and Motor Skills*, 88, 1235-1244.
- McCann, R., & Plunkett, R. P. (1984). Improving the concurrent validity of the Bender - Gestalt Test. *Perceptual and Motor Skills*, 58, 947-950.
- McKay, M. F., & Neale, M. D. (1985). Predicting early school achievement in reading and handwriting using major "error" categories from the Bender-Gestalt Test for young children. *Perceptual and Motor Skills*, 60, 647-654.
- Moose, D., & Brannigan, G. G. (1997). Comparison of preschool children's scores on the modified version of the Bender-Gestalt Test and the Developmental test of Visual-Motor Integration. *Perceptual and Motor Skills*, 85 (2), 766-775.

- Moura, M. L. S. (1982). Reconhecimento e reprodução de figuras do Teste Gestáltico Visomotor de Bender: Um estudo evolutivo. *Arquivos Brasileiros de Psicologia*, 34 (4), 50-61.
- Muñiz, J. (2004). *La validación de los tests*. Metodología de las Ciencias del Comportamiento. V5 (2), 121-141.
- Neale, M. D., & McKay, M. F. (1985). Scoring the Bender Gestalt Test using the Koppitz developmental system: Interrater reliability, item difficulty, and scoring implications. *Perceptual and Motor Skills*, 60 (2), 627-636.
- Neri, M. L. (2005). *Bender – Sistema de Pontuação Gradual: um estudo com crianças surdas*. Dissertação de Mestrado da Universidade São Francisco, Itatiba, São Paulo.
- Nielsen, S., & Sapp, G. L. (1991). Bender Gestalt developmental scores: Predicting reading and mathematics achievement. *Psychological Reports*, 69,39-42.
- Noronha, A.P.P., & Vendramini, C. M. M. (2003). Parâmetros Psicométricos: Estudo Comparativo entre Testes de Inteligência e de Personalidade. *Psicologia: Reflexão e Crítica*, 16 (1) pp. 177-182.
- Noronha, A. P. P. (2002). Instrumentos Psicológicos mais conhecidos por estudantes do Sul de Minas Gerais. *Avaliação Psicológica*, Vol 1 (2), 151-158.
- 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*, Vol 7 (1), 47-56.
- Pacheco, L. M. B., & Sisto, F. F. (2003). Ajustamento social e dificuldade de aprendizagem. *PSIC-Revista de Psicologia da Vetor Editora*, 6 (1), 43-50.

- Pascal, G. R., & Suttell, B. (1951). *The Bender Gestalt-test* quantification and validity for adults. New York: Grune & Stratton.
- Pasquali, L. (2001). *Técnicas de Exame Psicológico – TEP: Manual*. São Paulo: Casa do Psicólogo/ Conselho Federal de Psicologia.
- Pasquali, L. (2003). *Psicometria: teoria dos testes na psicologia e na educação*, Petrópolis RJ: Vozes.
- Pereira, A. P. C., & Carellos, S. D. M. (1995). Examinado o ensino das Técnicas de Exame Psicológico. *Cadernos de Psicologia*, 3 (4), 33-36.
- Pinelli Jr., B., & Pasquali, L. (1990). Parâmetros Psicométricos do Teste Gestáltico Visomotor de Bender: um estudo empírico. *Revista de Psicologia*, 1/2 (7/8), 51-74.
- Pinelli Jr., B., & Pasquali, L. (1991/1992). Parâmetros psicométricos do Teste Gestáltico Vismotor de Bender: um estudo empírico. *Revista de Psicologia*, 9/10 (1/2), 51-74.
- Pinelli, Jr., B., & Frey, P. D. (1991). Visual discrimination and visuomotor integration among two classes of brasilian children. *Perceptual and Motor Skill*, 72, 847-850.
- Posada, M. C. (2002). Um sistema alternativo para la evaluacion del Test Gestáltico de Bender: *Revista Ibero-americana de Diagnóstico y Evaluacion-Psicologica*, 14 (2), 9-31.
- Puente, M., & Maciel Jr., J. A. (1984). Procedimentos Multimodais no uso do Teste de Bender. *Estudos de Psicologia*, 1 (2), 65-74.
- Prieto, G., & Muniz, (2001). J. Um modelo para evaluar la calidad de los tests utilizados em Espana. Disponível em [http: www.cop.es/tests/modelohtm](http://www.cop.es/tests/modelohtm). Consulta feita em 15/09/2005

- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danmarks Paedagogiske Institute.
- Rie, E. D., Rie, H. E., Stewart, S., & Rettemnier, S.C. (1978). An analysis of neurological soft signs in children with learning problems. *Brain and Language*, 6(1), 32-46.
- Salvia, J.; & Ysseldyke, J. (1991). *Avaliação em educação especial e corretiva*. São Paulo: Manole.
- Santos, A. A. A (2005). Evidências de validade de critério para o Teste de Cloze. *Manuscrito não digitado. Programa de pos Graduação Stricto Sensu em Psicologia. Universidade São Francisco, São Paulo.*
- 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.
- Santos, A. A. A. (2004). O Cloze como Técnica de Diagnóstico e Remediação da Compreensão em Leitura. *Interação em Psicologia*, 8(2), 217-226.
- Santos, R. S. (2006). *Avaliação da maturidade percepto-motora e traços de personalidade*. Dissertação de Mestrado, Programa de Estudos Pós Graduados em Psicologia. USF, Itatiba, SP.
- Santos, A. A. A., Rueda, F. J. M., & Bartolomeu, D. (2006). Avaliação dos aspectos afetivos envolvidos nas dificuldades de aprendizagem. Em F. F. Sisto, & S. C. Martinelli (Orgs.), *Afetividade e Dificuldades de Aprendizagem* (pp.93-110). São Paulo: Vetor.

- Santucci, H., & Galifret-Granjon, N. (1968). Prova Gráfica de Organização Perceptiva. In R. Zazzo (Org.), *Manual para o Exame Psicológico da Criança* (pp.233-268). São Paulo: Mestre Jou.
- Santucci, H., & Pêcheux, M.G. (1981). Prova Gráfica de Organização Perceptiva para Crianças de 6 a 14 anos. In R. Zazzo (Org), *Manual para o Exame Psicológico da Criança* (pp.291-338). São Paulo: Mestre Jou.
- Shapiro, S. K., & Simpson, R. G. (1995). Koppitz scoring system as a measure of Bender-Gestalt performance in behaviorally disturbed adolescents. *Journal of Clinical Psychology*, 51(1), 108-112.
- Silva, M.J.M. (2002). *Avaliação da compreensão da leitura e desempenho acadêmico em universitários*. Dissertação de Mestrado. USF, Itatiba, SP.
- Sisto, F. F., Codenotti, N., Costa, C. A. J., & Nascimento, T.C.N. (1979). *Testes psicológicos no Brasil: que medem realmente*. *Educação e Sociedade*, 2, 152-165.
- Sisto, F. F., Noronha, A. P. P., & Santos, A. A. A. (2003). *Distorção de forma no Teste de Bender: Questionando seu critério de validade*. Manuscrito submetido à publicação.
- Sisto, F. F. (2001). *Dificuldade de Aprendizagem em Escrita: um instrumento de avaliação (ADAPE)*. Em: Sisto, F. F.; E. Boruchovitch, E.; Fini, L. D. T.; Brenelli, R. P. & Martinelli, S. de C. (Orgs), *Dificuldades de aprendizagem no contexto psicopedagógico*. Petrópolis: Vozes.
- Sisto, F. F.; Bueno, J. M. H.; & Rueda, F. J. M. (2003). Traços de Personalidade e Distorção e Integração de Formas: *Um estudo de validade*. *Psicologia em Estudo*, 8 (1), 77-84.

- Sisto, F. F. & Fernandes, D. C. (2004). Dificuldades Lingüísticas na Aquisição da Escrita e Agressividade. *Psicologia Escolar e Educacional*, 8 (1), 75-84.
- Sisto, F. F., Noronha, A.P.P., & Santos, A. A. A. (2004). Distorção de Forma no Teste de Bender: questionado seu critério de validade. *Revista do Departamento de Psicologia da UFF*, 16(2), 139-154.
- Sisto, F. F., Santos, A. A. A., & Noronha, A. P. P. (2004). Critério de integração do Teste de Bender: explorando evidências de validade. *Avaliação psicológica, São Paulo*, 3(1) 13-20.
- Sisto, F. F., Noronha, A. P. P & Santos, A. A. A. (2005). Manual Bender – Sistema de Pontuação Gradual (B-SPG). São Paulo: *Vetor Editora Psicopedagógica Ltda.*
- Sisto, F. F. (2005). *DFH-Escala Sisto*. São Paulo: Vetor Editora.
- Sisto, F. F., Martinelli, S. C. (2006). Afetividade e dificuldades de aprendizagem: uma abordagem psicopedagógica. São Paulo: *Vetor Editora Psicopedagógica Ltda.*
- Suehiro, A. C. B. (2005). *Contribuição ao Sistema de Pontuação Gradual de Bender: evidências de validade na escolarização*. Dissertação de Mestrado da Universidade São Francisco, Itatiba, São Paulo.
- Taylor, R. L., Kauffman, D., & Partenio, I. (1984). The Koppitz Developmental Scoring System for the Bender-Gestalt: Is it developmental? *Psychology in the-Schools*, 21(4), 425-428.
- Taylor, W. L. (1957). Cloze procedure: a new tool for measuring readability. *Journalism Quarterly*, 30,415-433.

- Tolor, A., & Schulberg, H. (1963). An Evaluation of the Bender Gestalt Test. Illinois, USA, Charles C. Thomas Publs.
- Weristein, E. S. (1983). Sex Differences in early Bender Gestalt Test performance. *Perceptual and Motor Skill*, 57, 301-302.
- Wertheimer, M. (1923). *Studies in the theory of Gestalt Psychology*, Psychol. Forsch., 4.
- Vance, B., Fuller, G. B., & Lester, M. L. (1986). A comparison of the Minnesota Perceptual Diagnost – Test Revised and the Bender-Gestalt. *Journal of Learning Disabilities*, V. 19 n°. 4. p.211-214.
- Zazzo, R. (1968). *Manual para Exame Psicológico da Criança* 1ª ed., São Paulo Ed. Mestre Jou.
- Zucolotto, K. A., & Sisto, F. S. (2002). Dificuldades de aprendizagem em escrita e compreensão em leitura. *Interação em Psicologia*, 6(2), 157-166.

VII – ANEXOS

ANEXO 1 – TESTE GESTÁLTICO VISOMOTOR DE BENDER

O teste pode ser de aplicação individual ou coletiva. No estudo em questão, a aplicação foi coletiva.

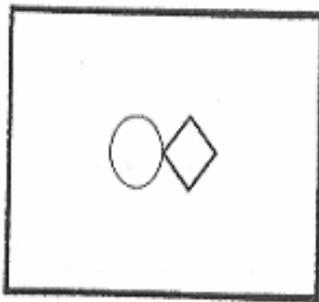
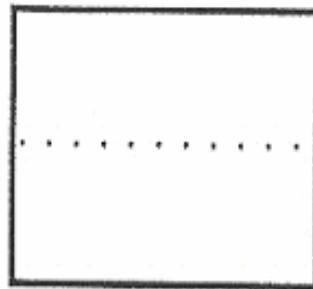
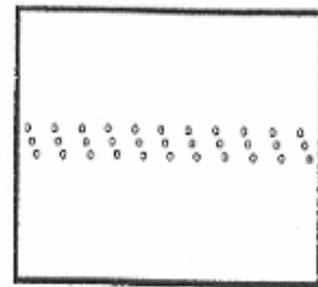
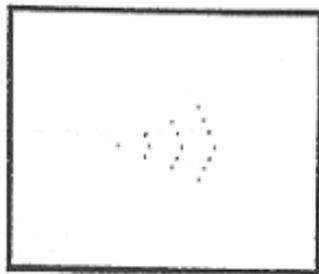
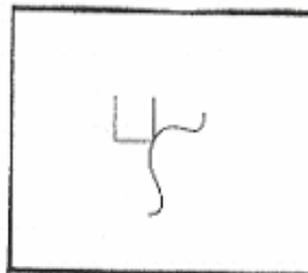
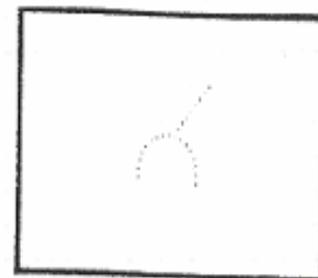
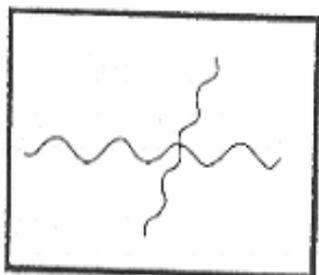
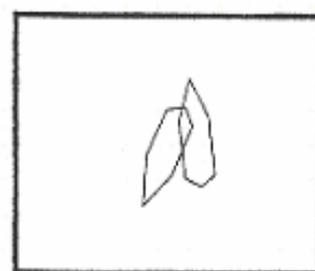
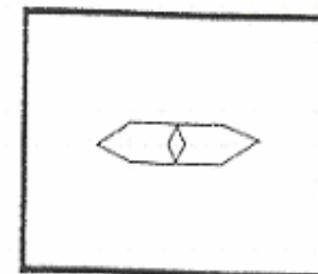
Instruções para aplicação do Teste Bender-SPG

-Explicar aos alunos que serão mostradas 9 figuras, as quais deverão ser desenhadas numa folha em branco.

-Reforçar a informação de que todas elas deverão ser copiadas em um lado só da folha.

-A instrução consiste em informar-lhe que será projetada uma figura por vez e que a transparência será mudada depois que todos os alunos tiverem terminado.

Seguem abaixo os modelos das Figuras A, 1, 2, 3, 4, 5, 6, 7, e a Figura 8 do Bender, a título de ilustração.

*Figura A**Figura 1**Figura 2**Figura 3**Figura 4**Figura 5**Figura 6**Figura 7**Figura 8*

ANEXO 2 – TESTE DE COMPREENSÃO EM LEITURA (CLOZE)

Instruções para aplicação do Teste de Cloze

- Solicitar aos alunos que leiam o texto completo;
- Depois, que escrevam a palavra que melhor complete o seu sentido;
- Lembrá-los de que o tamanho do traço corresponde ao tamanho da palavra que cabe em cada espaço.

A PRINCESA E O FANTASMA

Era uma vez uma princesa que vivia muito infeliz em seu palácio. Ela era apaixonada por um fantasma que vivia escondido lá.

Um dia chegou um misterioso estrangeiro e disse a ela que o seu fantasma era um príncipe enfeitiçado.

A princesa suspirou de alívio e ficou pensando em uma maneira de tirar aquele feitiço. Achou que se o fantasma soubesse do seu amor por ele, o feitiço desapareceria.

Acreditando nisso, a princesa armou um plano e prendeu o fantasma numa caixinha de música. Declarou seu amor a ele e, ao abrir a caixinha, o som da música o transformou num príncipe maravilhoso.

ANEXO 3 – TESTE DA ESCALA AVALIAÇÃO DE DIFICULDADES NA APRENDIZAGEM DA ESCRITA (ADAPE)

UMA TARDE NO CAMPO

José ficou bastante alegre quando lhe contaram sobre a festinha na chácara da Dona Vanda. Era o aniversário de Amparo.

Chegou o dia. Todos comeram, beberam e fizeram muitas brincadeiras engraçadas.

Seus companheiros Cássio, Márcio e Adão iam brincar com o burrico. As crianças gostam dos outros animais, mas não chegam perto do Jumbo, o cachorro do vizinho. Ele é mau e sai correndo atrás da gente.

Mário caiu jogando bola e machucou o joelho. O médico achou necessário passar mercúrio e colocou um esparadrapo.

Valter estava certo. Foi difícil voltar para casa, pois estava divertido.

Pensando em um dia quente de verão, tenho vontade de visitar meus velhos amigos.

Tabela 1 - Critério de Classificação de Dificuldade de Aprendizagem na Escrita do ADAPE para a segunda série.

Palavra errada	Categoria	2ª série
<i>Até 20 erros</i>	<i>0</i>	<i>Sem indícios de DA</i>
<i>50 a 79 erros</i>	<i>2</i>	<i>DA leve</i>
<i>80 ou + erros</i>	<i>3</i>	<i>DA média</i>

O intervalo verificado entre as crianças classificadas como ‘sem indícios de Dificuldade de Aprendizagem’ (até 20 erros) e as com ‘Dificuldade de Aprendizagem Leve’ (de 50 a 79 erros) apresenta crianças que, conforme Sisto (2001) estão na fase final da alfabetização e normalmente apresentam poucos erros. O autor reforça que é preciso considerar os problemas do sistema educacional vigente para não comprometer essas crianças, incluindo-as como portadoras de dificuldades de aprendizagem.

Tabela 2 - Critério de Classificação de Dificuldade de Aprendizagem na Escrita do ADAPE para a terceira série.

Palavra errada	Categoria	3ª série
<i>Até 10 erros</i>	<i>1A</i>	<i>Sem indícios de DA</i>
<i>11- 19 erros</i>	<i>1B</i>	<i>DA leve</i>
<i>20-49 erros</i>	<i>3</i>	<i>DA média</i>
<i>50 ou + erros</i>	<i>4</i>	<i>DA acentuada</i>

ANEXO 4 – TESTE DO RECONHECIMENTO DE PALAVRAS

Instruções

Prezada Professora,

Por favor, leia apenas uma vez cada palavra e dê o tempo necessário para a criança marcar a resposta certa.

Explicar para a criança o que ela terá de fazer, da seguinte forma: - Vocês vão receber uma folha com palavras. Eu vou ler a primeira palavra e vocês marcarão a forma correta de escrever a palavra que eu falei. Escrever o seguinte exemplo na lousa.

1. abacaxi – bacaxi – abacachi

- Eu falo a palavra abacaxi (falar claramente, sem silabar) e vocês fazem um X na resposta que acham que está correta. Assim (marcar na lousa).

- Mesmo que não tenham certeza, não deixem de marcar a palavra que vocês acham que está correta. Prestem atenção, pois eu não vou repetir a palavra.

Folha do aplicador

1- além	20- entrar	39- painel
2- enfim	21- errado	40- parece
3- gol	22- estiagem	41- pediátrico
4- palavra	23- exemplo	42- pegá-lo
5- vento	24- existe	43- propagandas
6- populações	25- exploração	44- proparoxítonas
7- posição	26- foram	45- receita
8- produz	27- forma	46- responsável
9- pronunciadas	28- futebol	47- socou
10- amarela	29- geralmente	48- surgiu
11- ambiente	30- imaginação	49- telas
12- atrasado	31- importante	50- viria
13- cinzenta	32- informação	51- vivacidade
14- clareira	33- juntaram	52- vontade
15- claro	34- lápis	53- pequenas
16- dela	35- madrugada	54- personagens
17- derradeiro	36- material	55- pesquisadas
18- descubra	37- médico	
19- distância	38- ninguém	

Teste do Reconhecimento de Palavras: folha do aluno

Escola:	Série:	Sexo:	Idade:
----------------	---------------	--------------	---------------

- 1 – alem - alén - além
- 2 – emfim - enfin - enfim
- 3 – jol - gou - gol
- 4 – palavra - pelavra - pavavra
- 5 – ventu - vento - vênto
- 6 – populações - populassões - popuações
- 7 – pusição - posicão - posição
- 8 – poduz - produz - produs
- 9 – pronunciada – pronunciadas- pronunssiadas
- 10 – marela - amarrela - amarela
- 11 – ambienti - ambiente - ambiente
- 12 – atrasado - atlasado - atrasado
- 13 – cinzenda - cizenta - cinzenta
- 14 – clarera - clareira - crareira
- 15 – clarro - claro - craro
- 16 – déla - delá - dela
- 17 – dêrradeiro - deradeiro - derradeiro
- 18 – decubra - desçubra - descubra

- 19 – ditância - distânsia - distância
- 20 – emtrar - entar - entrar
- 21 – enrado - erado - errado
- 22 – eztiagem - esfiagem - estiagem
- 23 – exemplo - exemplo - ixemplo
- 24 – eziste - existe - existi
- 25 – exploração - esploração - exproração
- 26 – foran - poram - foram
- 27 – forma - forna - fôrma
- 28 – fotebol - futebou - futebol
- 29 – geraumente - geralmente - geralmenti
- 30 – imaginasão - imaginação - imajinação
- 31 – importante - inportante - importamte
- 32 – informação - informassão - infornação
- 33 – juntaram - juntáram - juntaran
- 34 – lapis - lápis - lápiz
- 35 – madrugaba - madrugada - nadrugada
- 36 – matirial - materiau - material
- 37 – medico - médica - médico
- 38 – ninguem - ninguém - ninguém
- 39 – paineu - painel - panel
- 40 – paresse - padece - parece

- 41 – pediátricu - pediático - pediátrico
- 42 – pega-lo - pagá-lo - pegá-lo
- 43 – proqagandas - propagamdas - propagandas
- 44 – proparoxítunas - proqaroxítonas - proparoxítonas
- 45 – raceita - resseita - receita
- 46 – reponsável - resposável - responsável
- 47 – socou - çocou - socol
- 48 – çurgiu - surjiu - surgiu
- 49 – telas - temas - télas
- 50 – viria - tiria - virria
- 51 – vivacidabe - vivaacidade - vivacidade
- 52 - vontadi - vomtade - vontade
- 53 - pequenas - peqenas - piquenas
- 54 - personagens - perçonagens - persomagens
- 55 - pesquisadas - tesquisadas - pesquisaba

ANEXO 5: TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Pesquisa: Estudo de Evidência de Validade do Bender

Eu, _____, RG. nº _____, residente e domiciliado à _____, no município de _____, UF: _____, abaixo assinado, responsável legal por _____, dou meu consentimento livre e esclarecido para que ele (a) participe como voluntário do projeto de pesquisa supracitado, sob a responsabilidade dos pesquisadores Dra. Acácia A. Angeli dos Santos, Dra. Ana Paula Porto Noronha e Dr. Fermino Fernandes Sisto, professores do Programa de Estudos Pós - Graduação em Psicologia, na cidade de Itatiba/SP.

Assinando este Termo de Consentimento estou ciente de que:

1. O objetivo do estudo é buscar Evidência de Validade de um novo sistema de aplicação e correção do Bender em alunos do ensino fundamental;
2. Durante o estudo serão utilizados como instrumentos de avaliação os testes de Bender, Cloze, Reconhecimento de Palavras e a Escala de Avaliação de dificuldade na aprendizagem da escrita (ADAPE).
3. A aplicação destes testes não implica em nenhum risco para a sua saúde física e mental;
4. Obtive todas as informações necessárias para decidir conscientemente sobre a sua participação na referida pesquisa;
5. Estou livre para interromper a qualquer momento sua participação na pesquisa;
6. Seus dados pessoais serão mantidos em sigilo e os resultados gerais obtidos por meio da pesquisa serão utilizados apenas para alcançar o objetivo do trabalho, exposto acima, incluídas suas publicações na literatura científica especializada;
7. Poderei contatar o Comitê de Ética em Pesquisa da Universidade São Francisco para apresentar recursos ou reclamações em relação à pesquisa pelo telefone (011) 4034-8442;
8. Poderei entrar em contato com os responsáveis pelo estudo, sempre que julgar necessário pelo telefone (011) 4534-8000;
9. Esse Termo de Consentimento é feito em duas vias, sendo que uma permanecerá em meu poder e a outra com o pesquisador responsável.

_____, _____ de _____ de 2005.

Assinatura de Responsável Legal