

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE PSICOLOGIA

CRISTIANO ESTEVES

Padronização brasileira do Teste Bender Koppitz-2
em crianças da cidade de São Paulo

SÃO PAULO
2015

CRISTIANO ESTEVES

**Padronização brasileira do Teste Bender Koppitz-2
em crianças da cidade de São Paulo**

Tese apresentada ao Instituto de Psicologia da Universidade de São Paulo – IPUSP, como parte dos requisitos para a obtenção do título de Doutor em Psicologia.

Área de Concentração: Psicologia Escolar e do Desenvolvimento Humano.

Orientadora: Profa. Dra. Irai Cristina Boccato Alves

Versão Corrigida

SÃO PAULO

2015

NÃO AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE
TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO.

Catálogo na publicação
Biblioteca Dante Moreira Leite
Instituto de Psicologia da Universidade de São Paulo

Esteves, Cristiano.

Padronização brasileira do teste *Bender Koppitz-2* em crianças da cidade de São Paulo/ Cristiano Esteves; orientadora Irai Cristina Boccato Alves. -- São Paulo, 2015.

188 f.

Tese (Doutorado – Programa de Pós-Graduação em Psicologia. Área de Concentração: Psicologia Escolar e do Desenvolvimento Humano) – Instituto de Psicologia da Universidade de São Paulo.

1. Teste Gestáltico de *Bender* 2. Escala *Koppitz, Koppitz-2* 3. Distúrbios de aprendizagem 4. Validade do teste 5. Precisão do teste 6. Padronização do teste 7. Integração visomotora 8. Maturidade percepto-motora I. Título.

BF698.8.B4

Nome: Cristiano Esteves

Título: Padronização brasileira do Teste Bender Koppitz-2 em crianças da cidade de São Paulo.

Tese apresentada ao Instituto de Psicologia da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de Doutor em Psicologia.

Aprovado em: ____/____/____

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Assinatura: _____

AGRADECIMENTOS

A gratidão é um sentimento que nos liga com o Bem. É o reconhecimento de algo bom que recebemos. Uma força que nos impulsiona para seguir adiante. Neste momento da minha caminhada de Vida, encontro-me diante de um ciclo que se fecha para que outros se abram. Dessa forma é o momento de agradecer a todos que, de longe ou de perto, colaboraram com este trabalho.

À minha orientadora Prof^a Dr^a Irai Cristina Boccato Alves, obrigado por sua amizade, por dividir comigo seu conhecimento, possibilitando que eu me torne um profissional melhor. Sua paixão pela profissão e especificamente pela avaliação psicológica tomou conta de mim também, desde a época da graduação, no mestrado e agora no doutorado. Obrigado por sua paciência (muita paciência). Obrigado por acreditar em mim.

A Vetor Editora especialmente ao psicólogo José Glauco Bardella, suas filhas Marcia Boarini Bardella e Monica Bardella Garcia pelo apoio e colaboração de sempre em minha formação profissional. Ao Sr. Ricardo Mattos obrigado por sua compreensão, pelo apoio e incentivo para que eu levasse este projeto adiante. Obrigado por acreditar e investir em meu potencial.

À Prof^a. Dr^a. Helena Rinaldi Rosa obrigado por sua ajuda durante toda a pesquisa. Obrigado pelo apoio no contato com as escolas, na coordenação da equipe de aplicadores, na realização das aplicações e avaliações dos testes. Obrigado por me ouvir, pelo incentivo e por me estimular a continuar. Sua importância para a concretização desta pesquisa foi muito grande.

À Prof^a. Dr^a. Fraulein Vidigal de Paula por toda sua ajuda na coordenação do processo de coleta de dados e por me incentivar a fazer o doutorado. Foi difícil, mas valeu a pena.

À Prof^a. Dr^a. Adele de Miguel Marques por sua importante contribuição na coleta de dados. Meu sincero agradecimento.

Às psicólogas Tábata Cardoso, Anna Neves Lance, Emanuelle Arsuffi por disponibilizarem seu tempo de descanso para colaborar com a avaliação dos protocolos. Obrigado por todo apoio e incentivo, por sempre se colocarem disponíveis para ajudar.

À psicóloga Cristiane Raymundo Bezerra por sua ajuda na pesquisa e por me dar uns “empurrões” em momentos importantes. Sei como as coisas são corridas, mas mesmo

assim em suas horas livres encontrou uma forma de colaborar com esta etapa de minha vida. Obrigado acima de tudo por sua amizade.

Especialmente aos psicólogos Felipe Marangoni e Julia Maria Migot. Obrigado pela dedicação e disponibilidade para colaborar, de forma solícita atendendo inclusive a pedidos de maior urgência com os prazos. Obrigado pela colaboração, seja nas aplicações, seja nas avaliações dos testes. Foi muito bom conhecer vocês.

Obrigado à equipe de colaboradores que trabalharam na coleta de dados: Ana Carrelhas de Albuquerque, Júlia Maria Migot, Jaqueline Pizzitola, Mariana dos Santos, Jeny Kyung Sook Kim, Jéssica Ochoa, Plínio Oshiro e Felipe Marangoni. O trabalho de vocês foi muito importante para esta pesquisa.

Ao Prof. Ricardo Fatte e a Prof^a Dr^a Maria Isabel da Silva Leme, pela ajuda com o inglês.

À Ligia Furusawa psicóloga do LITEP, obrigado por colaborar com a coleta de dados. Além disso, obrigado por me incentivar e por sua preocupação comigo. Obrigado por sua amizade.

Obrigado às escolas, professores, diretores, orientadores pedagógicos por terem permitido a realização das coletas de dados.

Obrigado aos pais por permitirem que seus filhos colaborassem com esta pesquisa.

Às crianças o meu muito obrigado pela colaboração.

A minha família, meu pai minha mãe, irmã e afilhado, obrigado pelo apoio e compreensão. Vocês são muito importantes para mim.

Obrigado a todos que de uma forma ou de outra fizeram parte deste projeto e fazem parte da minha vida. Aos amigos do coração minha eterna gratidão por todas as horas e em especial pela força nos momentos mais complicados. Aos amigos de outros planos fica aqui meu agradecimento por toda ajuda.

Agradeço a Deus, senhor de tudo que existe fonte de amor universal pela oportunidade da Vida. Que eu possa aproveitá-la bem.

RESUMO

Esteves, C. (2015). *Padronização Brasileira do Teste Bender Koppitz-2 em crianças da cidade de São Paulo*. Tese de Doutorado, Instituto de Psicologia, Universidade de São Paulo, São Paulo. 188p.

O Teste Gestáltico Visomotor de Bender é um teste psicológico utilizado em psicodiagnósticos infantis nas áreas clínica e neuropsicológica, embora existam adaptações para adolescentes e adultos. Em crianças seu uso se destina a avaliações da maturidade percepto-motora ou organização visomotora, diagnósticos de lesões cerebrais e perturbações emocionais, podendo também ser utilizado como uma técnica projetiva. Para os adultos é empregado tanto como uma técnica projetiva quanto para a identificação de alguns indicadores psicopatológicos e neuropsicológicos. Desde a sua elaboração em 1938 por Lauretta Bender, diversos sistemas de aplicação e avaliação foram desenvolvidos, sendo um dos mais conhecidos o de Elizabeth Koppitz. A tarefa a ser realizada pelo examinando consiste na reprodução de alguns desenhos, que podem ser somente copiados ou copiados e posteriormente reproduzidos de memória, dependendo do sistema utilizado. A avaliação consiste na análise da forma e precisão com que os desenhos são realizados. Um dos mais novos sistemas de avaliação é o Teste Bender Koppitz-2 proposto por Reynolds, que não é somente uma revisão, mas também uma extensão e um redesenvolvimento do sistema Koppitz original e que foi desenvolvido com o objetivo de avaliar a capacidade de integração visomotora. A principal diferença em relação ao teste original é que são acrescentados novos modelos mais fáceis, para as crianças mais novas, e mais difíceis, para as mais velhas, além dos nove propostos por Bender. Também foi ampliada a faixa etária de utilização, que passou a ser de 5 a 85 anos. Além disso, enquanto na avaliação do teste original são pontuados os erros na reprodução dos desenhos, na versão atual são pontuados os acertos, ou seja, o quanto eles foram bem realizados. O objetivo deste trabalho foi realizar estudos de validade, fidedignidade e normatização do Sistema Koppitz-2 em crianças de escolas públicas da cidade de São Paulo. A amostra foi composta por 623 crianças, sendo 51,7% do sexo feminino e 48,3% do masculino. As idades variaram entre 6 e 12 anos (média= 9,05 anos e DP=1,86) e a escolaridade entre um e nove anos de estudo. A validade foi investigada em relação ao desenvolvimento e pela relação com outros testes, o Teste de Inteligência Não Verbal - R-2 e as Figuras Complexas de Rey e pela correlação com a versão original do teste. A fidedignidade foi estudada pelos métodos do teste e reteste, precisão entre avaliadores e consistência interna. Os testes foram aplicados individualmente e as crianças que participaram dos diferentes estudos foram incluídas na amostra total. As análises foram conduzidas separadamente para as crianças de 6 e 7 anos e para as de 8 a 12 anos, uma vez que os sistemas de pontuação baseados na qualidade da execução dos desenhos são diferentes em função da idade. Os resultados mostraram correlações entre 0,616 e 0,648 para os mais novos e entre

0,647 e 0,778 para os mais velhos com as Figuras Complexas de Rey. Em relação ao R-2 as correlações variaram entre 0,542 e 0,638 e 0,439 e 0,667 (respectivamente para os mais novos e mais velhos). As correlações com o sistema Koppitz foram de -0,646 para os mais novos e -0,566 para os mais velhos. Quanto à Fidedignidade, as correlações entre o teste e o reteste variaram entre 0,794 e 0,837 para os mais jovens e 0,745 e 0,964 para os mais velhos. Os valores do Alfa de Cronbach variaram entre 0,716 e 0,879 e para a precisão entre avaliadores as correlações foram de 0,966 para os mais novos e 0,972 para os mais velhos. Foram realizadas análises de variância em função da idade, sexo e tipo de escola, cujos resultados indicaram a necessidade de normas em percentil separadas em relação a essas três variáveis, já que as médias dos grupos se diferenciaram estatisticamente em relação a elas, com destaque para as diferenças entre as médias dos mais velhos que tenderam a ser maiores do que as dos mais novos, ainda que não em todos os casos. Os resultados indicaram adequados índices de validade e fidedignidade para o Koppitz-2 que confirmam suas propriedades psicométricas, caracterizando-o como um instrumento que pode ser utilizado nas avaliações da capacidade de integração visomotora de crianças de 6 a 12 anos nos psicodiagnósticos para os mais diversos fins. No entanto ainda serão necessárias algumas revisões nos critérios de avaliação, que em alguns casos, se mostraram pouco definidos.

Palavras-chave: Teste de Bender. Escala de Koppitz. Koppitz-2, Distúrbios de aprendizagem. Validade do teste. Precisão do teste. Padronização do teste. Integração visomotora. Maturidade percepto-motora.

ABSTRACT

Esteves, C. (2015). Brazilian standardization of the *Bender Koppitz -2 Test* in children from São Paulo city. Doctoral Thesis, Institute of Psychology, University of São Paulo, São Paulo. 188p.

The *Bender Gestalt Test* is a psychological test used in children diagnosis in clinical and neuropsychological areas, although there are adaptations for adolescents and adults. It is used to the assessments of perceptual-motor maturity or visual-motor organization, diagnosis of brain injuries and emotional disorders of children, and may also be used as a projective technique. For adults it is used both as a projective technique as well as for the identification of some psychopathological and neuropsychological indicators. Since it was published in 1938 by *Lauretta Bender*, many systems of administration and scoring have been developed, being one of the most known the system of Elizabeth Koppitz. The task to be performed by the examinee is to reproduce some drawings, which can only be copied or be copied and then reproduced by memory, depending on the system used. The evaluation consists in the analysis of the shape and the accuracy in which the drawings are done. One of the newest assessment systems is the *Bender Koppitz-2 Test*, proposed by Reynolds, which is not only a review, but also an extension and redevelopment of the original Koppitz system that was developed with the purpose of evaluate the visual-motor integration ability. The main difference is that new easier models for youngest children are added in the new test, and more difficult for the older ones, besides the nine, proposed by Bender. It has also been expanded the age range of use, from 5 to 85 years old. Furthermore, while in the evaluation of the original test errors are scored, in the current version the correct reproduction of the drawings are scored instead, meaning how well they were made. The goal of this work was to establish validity, reliability and the standardization of the Koppitz-2 system to children from public schools in the city of São Paulo. The sample consisted of 623 children, being 51.7% female and 48.3% male. The ages ranged from 6 to 12 years of age (average = 9.05 years, SD = 1.86) and education between one and nine years of study. The validity was investigated in relation to the development and by the relationship with other tests: The Nonverbal Intelligence Test R-2, the Rey Complex Figures and the original version of the test. The reliability was studied by the methods of test and retest, interscorers reliability and internal consistency. The tests were applied individually and children who participated in the different studies were included in the total sample. The analyses were made separately for children of 6 and 7 years old and from 8 to 12 years old since the scoring systems based on the quality of the drawings are different depending on the age. The results showed correlations between 0.616 and 0.648 for the youngest ones and between 0.647 and 0.778 for oldest ones with the Rey Complex Figure. Regarding the R-2, correlations ranged between 0.542 and 0.638 and 0.439 and 0.667 (respectively for the youngest and oldest children). The correlations with the original Koppitz system were -0.646 for the youngest ones and -0.566 for the oldest ones. Reliability coefficients between test

and retest varied between 0.794 and 0.837 for youngest and 0.745 and 0.964 for the oldest. The Cronbach's alpha ranged between 0.716 and 0.879 and the interscorers reliability correlations were 0.966 for the youngest ones and 0.972 for the oldest ones. Analysis of variance were performed according to age, sex and type of school, since their results indicated the need for standards in separate percentile to these three variables, and the average of the groups differed statistically in relation to them, especially the differences between the means of the oldest, who tended to be higher than of the youngest, but not in all cases. The results indicated adequate validity and reliability indices for the Koppitz-2, that confirm its psychometric properties, characterizing it as a tool that can be used in assessment of visual motor integration ability of children between 6-12 years of age in psychodiagnosis for various purposes. However some revisions will be required in the assessment criteria which, in some cases, proved to be poorly defined.

Key words: Bender Gestalt Test. Koppitz Scale. Koppitz-2. Learning disabilities. Test validity. Test reliability. Test standardization. Visual motor integration. Perceptual-motor maturity.

INFORMAÇÕES IMPORTANTES

A Lei 4119, de 27 de agosto de 1962 regulamenta a profissão de Psicólogo e no artigo 13º, 1º parágrafo dispõe que: **“Constitui função privativa do psicólogo o uso de métodos e técnicas para diagnóstico psicológico”** e, portanto, não pode ser exercida por outras pessoas sem essa formação e que não estejam inscritas no Conselho Regional de Psicologia.

O Código de Ética Profissional, publicado em agosto de 2005, estabelece as responsabilidades do Psicólogo:

- Artigo 1º (deveres fundamentais dos Psicólogos), Alínea i, estabelece que cabe a todo psicólogo **“zelar para que a comercialização, aquisição, doação, empréstimo, guarda e forma de divulgação do material privativo do psicólogo, sejam feitas conforme os princípios deste código”**;
- Artigo 2º estabelecendo o que é vedado ao Psicólogo, conforme alínea h, impede o profissional de **“interferir na validade e fidedignidade de instrumentos e técnicas psicológicas, adulterar seus resultados ou fazer declarações falsas”**;
- Artigo 15º determina que o Psicólogo deverá **“zelar pelo destino de seus arquivos pessoais, inclusive todo material psicológico (testes)”** e,
- Artigo 18º estabelece que: **“o Psicólogo não divulgará, ensinará, cederá, emprestará ou venderá a leigos instrumentos e técnicas psicológicas, que permitam e facilitem o exercício ilegal da profissão”**.

Diante do exposto, decidiu-se que a presente versão fosse reduzida para ser disponibilizada na internet. Tal redução foi necessária, pois houve a supressão de informações que possibilitassem a leigos compreenderem aspectos da aplicação ou interpretação do Teste Bender Koppitz-2.

Informamos que a versão completa da Tese, disponível para Psicólogos ou estudantes de Psicologia, encontra-se na biblioteca do Instituto de Psicologia da Universidade de São Paulo ou diretamente com o autor.

Cristiano Esteves

LISTA DE FIGURAS

Figura 1 - Histograma da distribuição do total de pontos do Koppitz-2 para o grupo dos mais novos.....	83
Figura 2 - Médias das pontuações por idade para as crianças de 6 e 7 anos	85
Figura 3 - Histograma da distribuição do total de pontos para o grupo de 8 a 12 anos	87
Figura 4 - Médias de pontos por idade para 8 a 12 anos	90
Figura 5 - Comparação do percentil 50 das normas americana e brasileira	93
Figura 6 - Frequência percentual de pontos dos itens por idade e para a amostra total de 6 e 7 anos	114
Figura 7 - Frequência percentual dos pontos dos itens do Koppitz- 2 por idade para o grupo de 8 a 12 anos	119

LISTA DE TABELAS

Tabela 10 - Distribuição de frequência da idade em função do sexo e da amostra total	66
Tabela 11 - Distribuição da escolaridade em função do sexo e da amostra total	67
Tabela 12 - Distribuição de frequência absoluta e percentual da escolaridade em função da idade	67
Tabela 13 - Distribuição de frequência da idade e sexo em função do tipo de escola	68
Tabela 14 - Distribuição de frequência por idade e sexo da amostra de validade com as Figuras Complexas de Rey	69
Tabela 15 - Distribuição por escolaridade e sexo da amostra de validade com as Figuras Complexas de Rey	69
Tabela 16 - Distribuição de frequência por idade e sexo da amostra de validade do R-2	70
Tabela 17 - Distribuição por escolaridade e sexo da amostra de validade com o R-2	70
Tabela 18 - Distribuição de frequência em função da idade e sexo da amostra do Reteste	71
Tabela 19 - Distribuição por escolaridade e sexo da amostra do Reteste	72
Tabela 20 - Tabela 20. Distribuição de frequência em função da idade e sexo da amostra de juízes	72
Tabela 21 - Distribuição da escolaridade em função do ano escolar e sexo da amostra de juízes	73
Tabela 22 - Distribuição de frequência em função da idade e sexo da amostra Koppitz (1963) e Koppitz-2	73
Tabela 23 - Distribuição da escolaridade em função do sexo da amostra Koppitz (1963) e Koppitz-2	74
Tabela 24 - Análise de Variância do total de pontos do Koppitz-2 por idade, sexo e tipo de escola para 6 e 7 anos	84
Tabela 25 - Médias e desvios padrão do total de pontos do Koppitz-2 dos mais novos	85
Tabela 26 - Análise de Variância do tempo de aplicação do Koppitz-2 por idade e sexo para os mais novos	86
Tabela 27 - Médias e desvios padrão do tempo de aplicação do Koppitz-2 dos mais novos por idade e sexo	86
Tabela 28 - Análise de Variância para o total de pontos do Koppitz-2 por idade, sexo e tipo de escola para 8 a 12 anos	88
Tabela 29 - Médias e desvios padrão do total de pontos do Koppitz-2 para 8 a 12 anos	88

Tabela 30 - Médias e desvios padrão do total de pontos do Koppitz-2 dos mais velhos em função da idade por sexo e tipo de escola.....	89
Tabela 31 - Resumo do Teste de Tukey por idade para os pontos do Koppitz-2 para 8 a 12 anos	90
Tabela 32 - Médias, desvios padrão e teste t por idade para o grupo feminino de crianças de 8 a 12 anos	91
Tabela 33 - Médias, desvios padrão e teste t por idade para o grupo masculino de crianças de 8 a 12 anos	91
Tabela 34 - Análise de Variância para o tempo de aplicação por idade, sexo e tipo de escola para 8 a 12 anos	92
Tabela 35 - Médias e desvios padrão do tempo de aplicação do Koppitz-2 para crianças de 8 a 12 anos	92
Tabela 36 - Correlações (r) entre os pontos da Figura de Rey B e Koppitz-2 para os mais novos por idade e por sexo	95
Tabela 37 - Correlações (r) entre a Figura A e Koppitz-2 para os mais velhos por idade e sexo	96
Tabela 38 - Correlações (r) entre o R-2 e Koppitz-2 para os mais novos	97
Tabela 39 - Correlações entre o R-2 e Koppitz-2 para os participantes mais velhos	98
Tabela 40 - Correlações entre os sistemas Koppitz-2 e Koppitz (1963/1989) por idade	100
Tabela 41 - Correlações entre os sistemas Koppitz-2 e Koppitz (1963/1989) por faixa etária e sexo para todas as figuras por idade ou apenas para as nove comuns	100
Tabela 42 - Correlações (r) para amostra total por idade e sexo dos mais novos – grupo do reteste	102
Tabela 43 - Correlações (r) para amostra de 10 e 11 anos por idade e sexo – grupo reteste ..	103
Tabela 44 - Alfa de Cronbach para o Koppitz-2 para 6 e 7 anos por idade e sexo	105
Tabela 45 - Alfa de Cronbach para o Koppitz-2 por idade, sexo e tipo de escola para 8 a 12 anos	105
Tabela 46 - Índices Kappa para os itens do Koppitz-2 para crianças entre 6 e 7 anos	107
Tabela 47 - Índices Kappa para os itens do Koppitz-2 crianças entre 8 e 12 anos	108
Tabela 48 - Correlação Item-Total (r_{it}) por sexo, idade e para amostra total para as idades de 6 e 7 anos	110
Tabela 49 - Variações dos alfas com a exclusão de itens com baixa correlação item-total grupo de 6 e 7 anos	111

Tabela 50 - Frequência percentual de acerto dos itens e qui-quadrado por idade para amostra de 6 e 7 anos	112
Tabela 51 - Correlação Item Total (r_{it}) por idade e para amostra total de 8 a 12 anos	115
Tabela 52 - Variações nos valores de alfa com a exclusão de itens com baixa correlação item-total r_{it}	116
Tabela 53 - Frequências percentuais de acerto dos itens e qui-quadrado por idade para crianças de 8 a 12 anos	117
Tabela 60 - Correlações por idade nas diferentes pesquisas sobre a relação entre o Bender e a Inteligência	131
Tabela 61 - Correlações por sexo nas diferentes pesquisas	132

SUMÁRIO

	Pág
I. INTRODUÇÃO.....	16
1 – O Teste de Gestáltico Visomotor de Bender	16
2- Sistemas de Aplicação e Avaliação do Bender	21
2.1 – O Teste Gestáltico de Bender para crianças de Elizabeth Koppitz	21
2.2 – A Prova Gráfica de Organização Perceptiva: o sistema Santucci-Granjon	27
2.3 – O Teste de Bender no sistema Santucci-Pêcheux	30
2.4 – O Teste de Bender como indicador de distúrbios emocionais em crianças: a avaliação Clawson	32
2.5 – Adaptação Hutt do Teste Gestáltico Visomotor de Bender para adultos	33
2.6 – Teste Gestáltico Visomotor de Bender - Sistema de Pontuação Gradual (B-SPG)	35
2.7 – Teste Gestáltico Visomotor de Bender – Segunda Edição (Bender Gestalt II) de Brannigan e Decker	37
2.8 – Sistema de Avaliação do Desenvolvimento de Koppitz para o Teste Gestáltico de Bender – Segunda Edição (Koppitz-2)	42
2.9 – Outros Sistemas de Avaliação	47
3 – Revisão da literatura: pesquisas com o Bender II e o Koppitz-2	50
4 – Pesquisas sobre a relevância do uso do Bender como instrumento de avaliação psicológica	63
II. OBJETIVOS	64
1 – Objetivo Geral	64
2 – Objetivos Específicos	64
III. MÉTODO	66
1 – Participantes	66
2 – Instrumentos	74
2.1 – Termo de Consentimento Livre e Esclarecido	74
2.2 – Sistema Bender Koppitz-2	74
2.3 – Teste das Figuras Complexas de Rey	74
2.4 – Teste de Inteligência Não Verbal para Crianças – R-2.....	77
3 – Procedimento	78
4 – Tratamento Estatístico	80
IV. RESULTADOS	83
1 - Diferenças entre idade, sexo e tipo de escola	83
2 – Validade	93
2.1 - Validade com o Teste das Figuras Complexas de Rey	94

2.2 - Validade com o Teste R-2	97
2.3 - Relação entre os sistemas Koppitz-2 e o original	99
3 – Fidedignidade	101
3.1 - Método do Teste-Reteste	101
3.2 - Consistência interna	105
3.3 - Precisão entre avaliadores	106
4 - Análise de Itens	109
4.1 - Resultados para as crianças de 6 e 7 anos	109
4.2 - Resultados para as crianças entre 8 e 12 anos	115
V. DISCUSSÃO	120
VI. CONSIDERAÇÕES FINAIS.....	139
VII. REFERÊNCIAS BIBLIOGRÁFICAS.....	142
VIII. ANEXOS	152
Anexo I – Termo de Consentimento Livre e Esclarecido.....	152
Anexo II – Instruções de aplicação Teste Bender Koppitz-2	154
Anexo III – Folha de registro do examinador – 5 a 7 anos	155
Anexo IV – Folha de registro do examinador – 8 a 85 ou mais	158

I. INTRODUÇÃO

1 – O Teste de Gestáltico¹ Visomotor de Bender

O Teste Gestáltico Vismotor de Bender, ou simplesmente Bender², é um teste psicológico amplamente conhecido e utilizado em avaliações psicológicas de crianças (nas áreas clínica e neuropsicológica) e em psicodiagnósticos para diversos fins (Bender, 1955; Clawson, 1962/1992; Cunha, 2000a; Koppitz, 1963/1974), embora existam adaptações que mostram que o teste também pode ser aplicado a adolescentes e adultos (Cunha, 2000b; Nunes, Ferreira e Lopes, 2007; Silva & Nunes, 2007). Para o grupo infantil sua utilização destina-se principalmente a avaliar a maturação percepto-motora (Nunes et al., 2007; Koppitz, 1963/1974), diagnóstico de lesões cerebrais e perturbações emocionais (Koppitz, 1963/1974) e também como uma técnica projetiva (Clawson, 1962/1992). Já para os adultos, o instrumento é mais usado como uma técnica projetiva (Cunha, 2000b; Hutt, 1969; Nunes et al., 2007), mas também pode ser usado para identificar alguns indicadores psicopatológicos e neuropsicológicos (Brannigan & Decker, 2003; Hutt, 1969; Reynolds, 2007).

O Teste Gestáltico e Vismotor de Bender foi construído por Laretta Bender entre os anos de 1932 e 1938 (Bernstein, 1955a). Seguindo a orientação da Teoria da Gestalt, trata-se de um teste clínico no qual se pede ao examinando para copiar figuras, apresentadas em cartões. Posteriormente, as cópias são analisadas para avaliar como os estímulos foram estruturados perceptualmente (Bernstein, 1955a).

Bender (1955) definiu a função gestáltica como:

função do organismo integrado através da qual ele responde a uma determinada constelação de estímulos como um todo sendo a resposta propriamente dita uma constelação, um padrão, uma gestalt. Todos os processos integradores do sistema nervoso se produzem em constelações, padrões ou gestalts. Esta integração ocorre não por adição, subtração ou associação, mas por diferenciação, ou aumentando ou diminuindo a complexidade interna do padrão, na sua estrutura. Ao que parece, um organismo integrado

¹ Nesta tese as palavras Gestalt e Gestáltico serão grafadas mantendo sua origem e não como Guestáltico ou Guestalt como são traduzidas algumas vezes para o português.

² O teste será denominado simplesmente de Bender, nomenclatura mais utilizada no Brasil para se referir ao instrumento.

nunca responde de uma forma diferente. O quadro total do estímulo e o estado de integração do organismo determinam o padrão de resposta. (p. 24)

Partindo desta concepção, é possível estudar a função gestáltica e as diversas condições integradoras e patológicas a ela envolvida a partir de um grupo de estímulos com estrutura mais ou menos semelhante. Apesar de afirmar que qualquer padrão do campo sensorial pode ser utilizado como um estímulo, a autora julga que os padrões visomotores são mais adequados, pois o campo visual se adapta mais facilmente a estudos experimentais e também facilita a cooperação dos examinandos/pacientes que devem apenas copiar alguns desenhos (Bender, 1955).

O teste foi elaborado tendo como ponto de partida 30 desenhos que foram desenvolvidos por Wertheimer em 1923, quando estudava a percepção visual. Nesta época, Lauretta Bender analisava desenhos infantis com a finalidade de avaliar a inteligência pelo nível de maturação neurológica observado nas reproduções gráficas (Nunes et al., 2007). Partindo destes estudos, ela selecionou nove figuras que foram consideradas importantes para “*avaliar a forma como o indivíduo percebe (visopercepto) e reproduz (motora) uma série de estímulos*” (Nunes et al., 2007, p. 42).

Dessa forma, pode-se dizer que a Psicologia da Percepção, tal como entendida pela Psicologia da Gestalt, serviu de base científica para o Teste de Bender. Para a elaboração deste instrumento a autora utilizou os princípios da Gestalt e o conceito de “*boa Gestalt*” de Wertheimer. Formulou uma série de conceitos e novos princípios a partir de suas investigações sobre a gênese da percepção da forma no desenvolvimento infantil, de algumas pesquisas sobre a Psicologia da criança normal e anormal publicadas por Kurt Koffka, Kurt Lewin, Heinz Werner e W. Wolff e das investigações de David Rapaport sobre o pensamento conceitual e as estruturas de emoção e da memória. Todo este trabalho colaborou de alguma maneira para a concepção do Teste Gestáltico de Bender (Bernstein, 1955a/2011).

O teste em sua forma original é composto por nove cartões com tamanho de 14,9 cm de comprimento e 10,1 cm de altura (Siva & Nunes, 2007), nos quais estão representados diversos desenhos (formas geométricas) constituídos por linhas retas e/ou curvas, linhas contínuas, pontos ou círculos. Os cartões são brancos e os

desenhos impressos em preto. A tarefa original do teste é copiar os estímulos dos cartões em uma folha de papel, contudo existem acréscimos na forma de aplicação como a reprodução das figuras “de memória” após a fase de cópia (Brannigan & Decker, 2003; Clawson, 1962/1992; Hutt, 1969; Reynolds, 2007) e fases de elaboração e associação (Clawson, 1962/1992; Hutt 1975, 1985). Os cartões apresentam numeração sequencial de um a oito que deve ser seguida durante a aplicação, sendo que o primeiro a ser copiado é chamado de Figura A, totalizando, portanto, nove cartões.

Silva e Nunes (2007) referem três princípios que regem a percepção dos estímulos visuais: fechamento, proximidade e continuidade. Segundo estas autoras, “em cada figura do Bender, algum destes princípios são enfocados, proporcionando entendimento quanto à estrutura e a forma do estímulo” (p. 78).

A Figura A é formada por um círculo e um quadrado desenhado com uma inclinação de 45° (na posição de um losango), sendo que um dos quatro cantos do quadrado toca o círculo. A Figura A foi escolhida como a primeira a ser desenhada, porque ela é rapidamente identificada como uma figura fechada sobre um fundo. Esta configuração formada por duas figuras contingentes é considerada uma “*gute Gestalt*” ou “*boa forma*” (Bender, 1955). Além deste princípio, pode-se identificar outro que também influencia a forma como a figura é percebida, o princípio de que as partes que estão mais próximas entre si tendem a ser visualizadas juntas. No caso da Figura A “*as partes contíguas do círculo e do quadrado estão mais próximas entre si do que dos lados do quadrado*” (Bender, 1955, p 24-25), mas na Figura A predomina o princípio do fechamento sobre o princípio da proximidade.

A Figura 1 normalmente é vista como uma linha de pontos, pois há uma tendência de fechar a gestalt, ou seja, a perceber a figura como um todo e não separada em partes (Silva & Nunes, 2007). No entanto, ela é formada por uma sequência de pares de pontos separados por uma mesma distância e os dois pontos das extremidades direita e esquerda não formam um par e deveriam parecer soltos. Por essa razão “*esses pares seriam percebidos com mais facilidade, se as diferenças na distância entre eles fossem maiores*” (Bender, 1955, p. 25). Esta figura é uma gestalt construída com base no princípio da proximidade das partes (Bender, 1955; Silva & Nunes, 2007).

A Figura 2 normalmente é percebida como uma sequência de linhas curtas inclinadas, formadas por três círculos (mais próximos) e não como três linhas

horizontais. Dessa forma, esta figura também é regida pelo princípio da proximidade das partes (Bender, 1955; Silva & Nunes, 2007).

Para a Figura 3 também se aplica o princípio da proximidade das partes (Bender, 1955; Silva & Nunes, 2007), pois nesta deve-se perceber linhas de pontos em forma de seta e não pontos separados entre si (Silva & Nunes, 2007).

A Figura 4 geralmente é vista como duas figuras inacabadas: um quadrado sem um dos lados, aberto na parte superior, e uma linha curva em forma de sino. Esta figura é marcada pelo princípio da continuidade da organização geométrica ou interna, já que o desenho é composto por duas figuras com formas geométricas distintas, como observado na Figura A (Bender, 1955; Silva & Nunes, 2007).

A Figura 5 é percebida como um círculo incompleto com um traço reto inclinado, ambos formados por linhas de pontos. Segundo Bender (1955/2011), o princípio que rege este figura é o mesmo que o da Figura A, ou seja, o princípio do fechamento. Para Silva e Nunes (2007) em função da maneira como esta figura pode ser produzida (círculo incompleto com traço) ela também pode ser percebida como continuidade.

A Figura 6 é formada por duas linhas sinusoidais de diferentes tamanhos (comprimentos de ondas) que se cortam obliquamente. Este cartão também segue o Princípio de Continuidade de Formas Geométricas.

As Figuras 7 e 8 também seguem o Princípio da Continuidade de Formas Geométricas e foram elaboradas usando as mesmas formas em posições diferentes, embora nem sempre isto seja percebido, especialmente no que diz respeito à Figura 7. Nessa figura, a posição das formas geométricas, que a constituem, ou seja, a figura da esquerda inclinada e a outra na posição vertical, além da inversão da orientação das figuras, colaboram para que as mesmas sejam percebidas como diferentes.

A Figura 7 é uma das mais difíceis para ser reproduzida por causa das mudanças nas direções das partes e dos ângulos que a compõem. Por este motivo a habilidade de copiar adequadamente esta figura é adquirida mais tardiamente pela criança.

A Figura 8 é constituída pelas mesmas duas formas que compõem a Figura 7, dispostas na posição horizontal lado a lado, sendo que o losango que se forma no centro, é fruto do cruzamento dos dois ângulos internos pela posição das figuras que estão em direções opostas. Esta figura é de execução aparentemente mais fácil,

quando comparada com a anterior, “*porque nela prevalece o Princípio da Continuidade da linha reta, sendo que nas pontas esquerda e direita da figura, o ângulo é mais aberto, formando duas pontas bem definidas*” (Silva & Nunes, 2007, p. 79).

Pode-se constatar, portanto, que cada uma das figuras foi elaborada seguindo um princípio específico que está vinculado à própria gestalt das figuras, mas todas envolvem a percepção visual (que é o foco de toda a estrutura do teste), a qual pode sofrer a modificações em sua forma ao ser processada pelo cérebro, se houver alguma deficiência nesse processo.

Para Nunes et al. (2007, p. 42), o objetivo original do Bender é:

observar de que forma surgem geneticamente as *gestalts* nas crianças e quais os processos de maturação envolvidos. A finalidade da autora foi entender que tipos de erros poderiam ocorrer na percepção de um estímulo dado (as figuras do teste) e se estes seriam decorrentes de distúrbios a nível cerebral ou de imaturidade para perceber e reproduzir corretamente.

Bender, no entanto, não chegou a desenvolver uma sistematização de correção do teste, ou seja, não elaborou um método objetivo de avaliação e pontuação dos desenhos (Nunes et al., 2007). Um de seus estudos, realizado com 800 crianças com idades entre 3 e 11 anos, teve como objetivo demonstrar a evolução dos desenhos em função do aumento da idade da criança. A partir deste estudo, a autora elaborou um quadro com as respostas mais comuns para cada faixa etária. Entre seus resultados ela pôde constatar a existência de uma relação entre a idade cronológica e a maturidade neurológica. Constatou ainda que aos 11 anos os desenhos eram realizados com uma quantidade muito reduzida de erros (Bender, 1955; Nunes et al., 2007).

Bender (1955) também usou o teste em avaliações clínicas em pacientes adultos com lesão cerebral, esquizofrenia, psicose e deficiência mental. Dessa forma enfatizou que o sujeito reage ao estímulo dado com um ato motor, conforme suas possibilidades maturacionais, uma vez que, algumas alterações observadas na percepção dos estímulos do teste em pacientes com lesões específicas em algumas partes do cérebro, demonstravam melhora na reprodução dos desenhos após tratamento neurológico (Nunes et al., 2007).

A revisão da literatura mostra que existe uma diversidade de métodos de aplicação, sistematização de avaliação e interpretação do Bender que foram desenvolvidos por diversos autores. Em relação à aplicação alguns autores administram somente a fase de cópia, em que se solicita que os desenhos sejam copiados (Koppitz, 1963/1989; Reynolds, 2007; Santucci & Galifret-Granjon, 1968; Sisto, Noronha & Santos, 2005). Outros aplicam também a fase de memória, que é executada após a cópia, e consiste em pedir que os desenhos sejam reproduzidos novamente sem ver os cartões (Brannigan & Decker, 2003; Clawson, 1962/1992; Hutt, 1975). Clawson (1962/1992) e Hutt (1975) propõem ainda que sejam aplicadas as fases de elaboração e de associação. Na elaboração pede-se que os desenhos sejam executados novamente com a orientação de que os mesmos sejam alterados e/ou combinados. Já na associação, não é realizado nenhum desenho, somente pergunta-se ao examinando com o que cada um deles se parece.

Os sistemas de avaliação podem ser classificados em função do grupo a que se destinam. Neste sentido podem ser voltados para avaliação infantil ou adulta. Na infantil podem ser citados os sistemas de Santucci e Galifret-Granjon (1968); Santucci e Pêcheux (1981); Koppitz (1963/1974); Clawson (1962/1992) e Sisto, Noronha e Santos (2005). Para os adultos o sistema mais conhecido é o proposto por Hutt (1969, 1975). Mais recentemente foram criados dois novos sistemas que visam atender ao público infantil, adolescentes, adultos e idosos: são as propostas de Brannigan e Decker (2003) e de Reynolds (2007), que além de ampliarem a faixa de utilização do instrumento, trabalham com 16 cartões, em que foram acrescentados desenhos mais simples e mais complexos aos originais para possibilitar uma avaliação mais precisa. A seguir, serão apresentados alguns destes sistemas.

2 – Sistemas de Aplicação e Avaliação do Bender

2.1 – O Teste Gestáltico de Bender para crianças de Elizabeth Koppitz

Koppitz (1963/1989) desenvolveu um sistema de pontuação do Bender com o objetivo de avaliar a maturidade da organização perceptivo-motora infantil. Suas pesquisas compararam a produção de crianças com diferentes idades o que permitiu estabelecer certos padrões de reprodução esperados em cada faixa etária. A partir

destes padrões, tornou-se possível identificar se a produção de uma determinada criança é adequada para sua idade ou semelhante à produção de uma criança com idade superior ou inferior a sua.

Além deste objetivo inicial, os estudos desta autora também permitiram a elaboração de uma lista de indicadores emocionais para crianças entre 5 e 10 anos, referentes à amostra por ela pesquisada (Cunha, 2000a; Domingues, Alves, Rosa & Sargiani, 2012; Koppitz 1963/1989; Nunes et al., 2007).

Partindo da observação de crianças, Koppitz (1963/1989) se interessou em determinar as relações que poderiam ser estabelecidas entre problemas emocionais, dificuldade de aprendizagem e a forma como são reproduzidos os desenhos do teste. Seus estudos a levaram a realizar a análise sistemática da produção das crianças de modo a poder verificar, se existiam diferenças na forma como elas reproduziam os desenhos ao longo dos anos. A partir desta observação foi possível levantar algumas hipóteses para explicar as diferenças na produção: elas poderiam estar ligadas a diferenças “*no tempo de maturação e na sequência com que elas aprendem as diversas funções gestálticas visomotoras*” (p. 20), mas também poderiam ser indicadores de problemas emocionais ou de lesões cerebrais.

A análise dos desenhos copiados permitiu constatar que algumas crianças conseguiam reproduzir a configuração da gestalt do desenho, mas tinham dificuldade em reproduzir alguns detalhes específicos; outras apesar de conseguirem reproduzir os detalhes, apresentavam dificuldades na orientação, integração e posição dos desenhos. Surgia então a necessidade de elaboração de uma escala que possibilitasse a avaliação do protocolo como um todo e não de cada desenho separadamente, a fim de identificar de forma mais precisa a maturidade perceptiva-motora da criança.

A escala maturacional foi elaborada a partir de uma lista de 20 alterações/distorções observadas nos protocolos de crianças. Após uma pesquisa inicial realizada com 77 crianças em idade escolar e que foram divididas em dois grupos, um com melhor rendimento e bom ajustamento e outro, com pior rendimento e dificuldades no ajustamento geral, as comparações mostraram que das 20 distorções apontadas inicialmente, apenas sete diferenciaram significativamente os grupos. São elas: distorção da forma, rotação, substituição de pontos por círculos, perseveração, integração das partes, ângulos ao invés de curvas e ângulos incorretos. Uma vez que estas sete categorias podem ser avaliadas em diversas

figuras do teste, foi elaborada a primeira versão da Escala de Indicadores Maturacionais de Koppitz formada por um total de 31 itens a serem avaliados.

Koppitz realizou diversos estudos com a finalidade de determinar as qualidades psicométricas da escala. Apresentou dois estudos de validade de critério, comparando os resultados entre grupos clínicos e o de controle. Um deles, já mencionado anteriormente, comparando os resultados do Bender de crianças com bom rendimento escolar e ajustamento com outras com baixo rendimento e dificuldades no ajustamento. O outro estudo foi feito com 51 pacientes (crianças) em atendimento em uma clínica, que foram divididos em dois grupos: pacientes com baixo rendimento escolar e problemas emocionais e pacientes com rendimento dentro da média e sérios problemas emocionais. Ambos os estudos mostraram a qualidade da escala. A precisão foi obtida pelo método de análise de juízes (concordância entre avaliadores).

A escala inicial passou por uma revisão em que foram ajustados alguns critérios de pontuação (especificamente em relação à perseveração na Figura 1 e na Figura 2). A nova escala, composta por 30 itens, foi submetida a novos estudos de validade (comparação entre grupos contrastantes tendo como critério o desempenho no Teste Metropolitano de Aproveitamento, dividindo a amostra em alto e baixo desempenho) e precisão (teste-reteste e análise de juízes) todos evidenciando resultados adequados.

Dessa forma pode-se dizer que a Escala de Indicadores Maturacionais de Koppitz se desenvolveu e foi validada, tendo como critérios o desempenho escolar, considerando que um aluno que apresente uma percepção visomotora bem desenvolvida provavelmente terá maior possibilidade de ter um bom desempenho, enquanto que aquele em que esta percepção ainda não foi amadurecida poderá ter dificuldades nas atividades escolares (Cunha, 2000a).

Na proposta de Koppitz (1963/1989) o teste deve ser aplicado individualmente, sendo necessárias duas folhas de papel, um lápis e uma borracha, que deve ficar disponível, mas a criança não deve ser estimulada a utilizá-la. Para a cópia a criança é informada de que deverá copiar os desenhos dos nove cartões. Estes permanecem sobre a mesa com a frente virada para baixo, sendo apresentados um de cada vez, à medida que forem concluídos os desenhos. Não há limite de tempo para a realização do teste, deve-se anotar eventuais comentários e comportamentos durante a aplicação. No caso de perguntas específicas sobre como

realizar a tarefa, deve-se responder de forma neutra, ressaltando a importância de fazer os desenhos o mais parecido possível com o modelo. Na Tabela 1 são apresentados os itens da Escala de Indicadores Maturacionais de Koppitz.

Os Indicadores Maturacionais de Koppitz se concentram em quatro categorias de desvios (Cunha, 2000a):

- a) Distorção de Forma: está relacionada aos aspectos formais pontos, linhas retas e/ou curvas e ângulos. Ocorre quando os aspectos da forma da figura são executados sem precisão: sem reproduzir corretamente os pontos, linhas retas, curvas e ângulos ou com substituição, omissão, multiplicação, deformação exagerada e perda da proporção relativa;
- b) Integração (ou desintegração, segundo Cunha, 2000a): quando há perda da configuração total ou parcial da figura seja pela separação entre as partes, omissão, acréscimo ou substituição de elementos que a constituam, perda da posição ou modificação estrutural acentuada na forma;
- c) Rotação: é a mudança na orientação do desenho em mais de 45°, quando comparado ao modelo. Se o desenho for feito na posição correta, mas o cartão for virado para a reprodução, também se considera rotação;
- d) Perseveração: refere-se ao aumento do número de elementos que compõem a figura original, sejam pontos, círculos ou curvas em relação ao modelo.

No sistema proposto por Koppitz, cada item recebe pontuação “0” ou “1” que corresponde, respectivamente, à ausência ou presença da característica. As pontuações são atribuídas às distorções/alterações no formato das figuras e dessa maneira, quanto maior a pontuação, pior é a qualidade da produção da criança e, portanto, maior poderá ser o déficit em sua percepção visomotora. A pontuação máxima é de 30 pontos. As pontuações somente são atribuídas se a alteração na forma da figura estiver claramente representada. No caso de dúvidas na atribuição de pontos, opta-se por não pontuar, pois os desvios menores devem ser ignorados, uma vez que a escala proposta foi desenvolvida para crianças que ainda não apresentam um controle motor totalmente desenvolvido (Koppitz, 1963/1989; Nunes et al., 2007).

Koppitz (1963/1989) elaborou normas para sua Escala de Indicadores Maturacionais com base em uma amostra de 1.104 crianças com idades entre 5 anos e 10 anos e 11 meses. A partir destes dados foi construída uma tabela com as médias e desvios padrão das crianças considerando faixas de idade com amplitude de 6 meses. Estes resultados permitiram estabelecer padrões para verificar se a pontuação atingida por uma determinada criança estava dentro do esperado para sua idade ou se o desempenho era semelhante ao de crianças com idade menor ou maior que a sua e assim avaliar sua maturidade perceptivo-motora. Uma pontuação maior do que a esperada para sua faixa de idade indica um déficit na percepção visomotora, enquanto que uma pontuação menor pode indicar uma maturidade típica de crianças mais velhas e, portanto, mais maduras.

Os dados de Koppitz mostraram que ocorre uma diminuição da média do total de pontos com o aumento da idade (a média de pontos para crianças de 5 a 5 anos e meio é de 12,1 enquanto que a média para a faixa de 10 a 10 anos e meio é de 3,1). Estes resultados são coerentes com a teoria que fundamenta o teste, pois uma das propostas do Teste de Bender é avaliar a maturidade visomotora que deve se desenvolver com o aumento da idade (inclusive do ponto de vista biológico). Dessa forma, a expectativa é que as distorções na reprodução dos desenhos diminuam com o aumento da idade, conforme as crianças se tornam mais maduras. Segundo Bender (1955), em geral aos 11 anos as crianças atingem a maturidade que permite que elas copiem os modelos sem erros.

Kroeff (1988, 1991) realizou uma pesquisa com 1.082 crianças de escolas públicas e privadas da cidade de Porto Alegre (RS), com idades entre 5 e 10 anos e 11 meses, para elaborar normas para o sistema de pontuação de Koppitz. As crianças foram agrupadas em função do tipo de escola e da idade, considerando o intervalo de 6 meses. Assim como no estudo de Koppitz, também foi observada uma diminuição na média do total de pontos com o aumento da idade. As médias das crianças das escolas públicas foram maiores que as das escolas particulares o que significa que o desempenho das crianças da escola pública foi pior (Cunha, 2000a). As médias da pesquisa de Kroeff (1988, 1991) foram maiores que as de Koppitz (1963/1989) com exceção dos resultados do grupo com idade entre 5 anos e 5 anos e meio. Estes resultados sugerem que as crianças brasileiras apresentaram maturidade percepto-motora menor que as americanas, pelo menos na época da pesquisa.

A partir da aplicação do teste em crianças com problemas emocionais e dificuldades de aprendizagem foi elaborado por Koppitz (1963/1989) outro sistema de avaliação: os indicadores emocionais para distinguir as distorções que estão relacionadas à imaturidade ou disfunção perceptiva daquelas que independem da idade e percepção, mas que refletem problemas emocionais.

Para a investigação dos indicadores emocionais a autora estudou 272 crianças entre 5 e 10 anos, sendo 136 encaminhadas a clínicas ou psicólogos, por apresentarem problemas emocionais e 136 alunos sem este tipo de queixa. Os grupos foram balanceados em função do sexo e da idade e os resultados mostraram que as crianças com problemas emocionais muitas vezes apresentavam imaturidade em sua percepção visomotora, sugerindo que eles podem interferir no processo de maturação. Entre os 11 indicadores emocionais avaliados, apenas seis diferenciaram de forma estatisticamente significativa os grupos de crianças com e sem problemas emocionais. Para as crianças de 5 a 7 anos foram significantes a presença de linha ondulada, linhas reforçadas e expansão. Já entre os participantes com idades entre 8 e 10 anos as diferenças significantes ocorreram para ordem confusa, tamanho pequeno, segunda tentativa de reprodução do desenho e expansão (Cunha 2000a; Koppitz, 1963/1989). Somente a expansão foi significativa para os dois grupos.

Para Koppitz (1963/1989) as perturbações emocionais são decorrentes dos problemas de percepção visomotora, ou seja:

as crianças com problemas na percepção visomotora experimentam muita frustração e frequentes fracassos, na sua casa e na escola, como resultado disto, muitas dessas crianças desenvolvem atitudes negativas e inadaptação emocional (p.154).

Koppitz (1963/1989) destaca ainda que há uma tendência para que crianças com percepção visomotora imatura venham a apresentar dificuldades de aprendizagem e problemas emocionais. No entanto não se pode generalizar que toda criança nesta condição irá necessariamente apresentar estes problemas. Recomenda ainda que sejam realizadas as duas análises, ou seja, dos indicadores maturacionais e dos emocionais. Para identificar problemas de percepção visomotora deve-se usar a escala maturacional e para avaliar as questões emocionais, devem ser usados os indicadores emocionais, que também podem ser

interpretados de forma projetiva. Segundo a autora, a presença de três ou mais indicadores é considerada superior à média e dois ou menos, inferior. Estes parâmetros de avaliação foram estabelecidos considerando que no grupo controle (sem problemas emocionais) o número máximo de indicadores encontrado foi quatro, enquanto que no grupo com problemas emocionais este índice chegou a seis. Outra observação realizada por Koppitz se refere à incidência dos indicadores que parece estar associada à gravidade de perturbação emocional:

mais da metade das crianças com três indicadores emocionais estavam emocionalmente perturbadas, quatro de cada cinco crianças com quatro indicadores emocionais em seus protocolos tinham problemas emocionais e todas as crianças com cinco ou mais indicadores emocionais haviam sido enviadas a psicólogos por sérios desajustes emocionais (Koppitz, 1963/1989, p.163).

2.2 – A Prova Gráfica de Organização Perceptiva: o sistema Santucci-Granjon.

Santucci e Galifret-Granjon (1968) apresentaram uma adaptação do Teste Bender para crianças entre 6 e 14 anos utilizando cinco das nove figuras do teste (Figuras A, 2, 3, 4 e 7). As autoras fizeram também uma alteração na ordem de aplicação em relação à original, invertendo a posição das figuras 3 e 4 para evitar que fiquem na sequência os estímulos formados por pontos. Neste sistema a avaliação é voltada para situações em que são observados problemas e/ou atrasos nos processos de aprendizagem escolar, que podem estar associados a “um atraso mental global” o que justifica “*examinar as possibilidades de uma criança no plano da organização perceptiva e motora no espaço*” (Santucci & Galifret-Granjon, 1968, p. 233). Para as autoras, a utilização do teste de Bender segundo sua proposta pode ajudar a identificar as discrepâncias entre o nível intelectual global da criança e seu nível de organização espacial, permitindo também analisar suas dificuldades.

Dessa forma fica claro que no sistema Santucci-Granjon (1968) não há a intenção de avaliar características de personalidade, ou seja, não são realizadas “interpretações caracterológicas”, pois para elas o teste avalia a organização perceptivo-motora do espaço que evolui com a idade e tem por objetivo examinar um aspecto: a realização gráfica. Quando esta organização é, por algum motivo,

prejudicada, isto tem um reflexo em várias atividades, entre elas, a aprendizagem escolar.

Inicialmente as autoras utilizavam o teste somente nos casos em que existia a hipótese de perturbações da organização espacial que estavam associadas a perturbações de linguagem. Começaram então um trabalho na área da Psicologia Escolar que culminou na definição de três aspectos que interferem na reprodução dos modelos e que foram utilizados no seu sistema: a construção dos ângulos; a orientação das figuras ou de seus elementos e a posição relativa das figuras entre si ou dos elementos que a compõem. Desenvolveram então um conjunto de critérios bastante exigentes, para avaliar estes três aspectos nos cinco modelos selecionados, em crianças entre 6 e 14 anos, tendo como foco estudar o progresso na qualidade de reprodução dos desenhos com o aumento da idade para verificar o quanto é possível chegar à perfeição na reprodução. Contudo, afirmam que esta não foi alcançada nem aos 11 anos como preconizava Bender (1955), nem aos 14 anos, idade limite de sua pesquisa.

Na adaptação de Santucci e Galifret-Granjon (1968) somente a cópia é solicitada, não é permitido o uso de borracha ou qualquer objeto que possa auxiliar na reprodução dos desenhos.

A avaliação considera três aspectos: a) os ângulos; b) a orientação dos elementos e c) a posição relativa dos elementos. Para cada um destes aspectos em cada desenho a criança pode receber uma pontuação que varia entre 0 e 3 pontos e quanto maior a pontuação, maior a qualidade da reprodução. São atribuídos ainda pontos adicionais para cada desenho, quando a cópia apresentar maior exatidão. Neste sistema a pontuação máxima é de 50 pontos, sendo no máximo 10 pontos para a Figura A (em que somente o quadrado é avaliado para ângulos); 8 pontos para a Figura 2 (neste modelo não são avaliados os ângulos); 11 pontos para a Figura 4; 11 pontos para a Figura 3 e 10 pontos para a Figura 5.

A pesquisa de Santucci e Galifret-Granjon (1968) foi realizada com uma amostra de 305 participantes de ambos os sexos com idades entre 6 e 14 anos. Para estudar a evolução da representação gráfica dos desenhos foram apresentados os dados em função da idade de três formas diferentes, considerando:

1 - A pontuação total: observou-se um aumento na pontuação com o aumento da idade principalmente até os 10 anos, mas que se estende até os 14, ainda

que em uma progressão menor o que, segundo as autoras, é causado pela exigência dos critérios por elas desenvolvidos;

2 – Os três aspectos avaliados separadamente: ângulos (onde foi observada uma regularidade na progressão da pontuação com o aumento da idade, sendo atingida a pontuação máxima aos 14 anos); orientação (que não apresentou uma regularidade no aumento, sendo o progresso na pontuação praticamente interrompido aos 10 anos e ainda que ocorra um leve aumento depois desta idade, não chega à pontuação máxima aos 14 anos); e a posição relativa (com resultados semelhantes aos ângulos, ainda que com uma pausa aos 10 anos).

3 – A frequência de pontuação em cada aspecto: para os ângulos há uma diminuição regular de pontuações 0 e 1 e o crescimento regular da pontuação com o aumento da idade; para a orientação também há um progresso (ainda que não tão regular) observado principalmente nas categorias extremas; para a posição relativa há um progresso muito claro entre os 6 e 10 anos, que depois se estabiliza.

As autoras apresentaram também os resultados por modelo, em que puderam verificar que a ordem de apresentação dos mesmos na aplicação não corresponde à da dificuldade da reprodução. Em uma comparação dos resultados em função do sexo dos participantes, constataram que a pontuação das meninas foi sistematicamente inferior, em relação à dos meninos. No entanto, destacam que estes resultados exigem estudos complementares para serem confirmados.

Foram feitos estudos psicométricos de precisão entre avaliadores e de validade relacionada ao desenvolvimento, que confirmaram a qualidade do instrumento para a avaliação principalmente de crianças entre os 6 e 10 anos, sendo que acima desta idade o método é mais sensível em situações associadas a quadros patológicos. São apresentadas ainda considerações sobre o cuidado ao se conduzir estudos clínicos, a importância da observação do comportamento, da utilização de outras técnicas de avaliação e eventuais interferências de aspectos relacionados à personalidade, que não pode ser deixada de lado, ainda que não seja o objetivo do sistema proposto.

Os estudos de Santucci e Galifret-Granjon (1968) devem, no entanto, ser considerados com certa cautela, pois o tamanho da amostra foi bastante reduzido (considerando o total de participantes e que nove grupos etários foram pesquisados, cada um ficou com cerca de 30 crianças) o que pode ter interferido significativamente nos resultados.

2.3 – O Teste de Bender no sistema Santucci-Pêcheux

Santucci e Pêcheux (1981), baseando-se em parte em sua experiência anterior com o Teste de Bender no sistema Santucci e Galifret-Granjon (1968), desenvolveram um novo sistema de avaliação mais objetivo (com menor interferência da subjetividade do avaliador) e mais rigoroso, destacando a importância do estudo genético da reprodução dos desenhos que permitisse realizar com mais clareza a discriminação da qualidade da produção em função da idade. Além disso, que possibilitasse também analisar as dificuldades da criança na idade escolar, no que diz respeito ao domínio grafoperceptivo.

Estudaram os resultados da população normal, a evolução da frequência das características entre as idades e o desaparecimento ou inexistência de certos tipos de reprodução conforme a idade. Para isso investigaram a organização grafoperceptiva e sua evolução no decorrer do avanço da idade escolar.

Sua pesquisa foi conduzida com 580 crianças, de ambos os sexos, com idades entre 6 e 14 anos de diferentes escolas de Paris e região. Assim como no sistema de Santucci e Galifret-Granjon (1968) foram aplicados apenas cinco desenhos do Bender original, sendo mantida a mesma ordem de aplicação proposta nesse sistema (ou seja, a ordem de aplicação dos desenhos 3 e 4 foi invertida). Foram incluídas algumas alterações na forma de aplicação: a folha é colocada na posição horizontal, a criança é informada sobre quantos desenhos irá realizar, indica-se o local da folha onde ela deve começar para que todos os desenhos possam ser feitos na mesma folha e o cartão com o desenho que será copiado fica sobre a folha e não ao lado como nos outros sistemas.

A avaliação é feita de forma bastante rigorosa, considerando sete critérios, embora nem todos se apliquem a todos os desenhos: 1 – Forma (aspectos relativos à forma do desenho e de suas partes); 2 – Número de elementos que compõem o desenho; 3 – Relação de proximidade (separação entre os elementos e/ou partes);

4 – Orientação Geral (mais primitiva, direita-esquerda, alta-baixa); 5 – Orientação Precisa (vertical, horizontal, oblíqua); 6 – Precisão dos alinhamentos e 7 – Exatidão das Dimensões.

A avaliação do sistema proposto por Santucci-Pêcheux exige a utilização de cinco transparências de correção, já que cada desenho tem seus próprios critérios de avaliação, ainda que estes possam ser agrupados nas categorias mencionadas anteriormente. As transparências são utilizadas para medir os ângulos, comprimentos, alinhamentos, paralelismos, inclinações em relação à vertical e à horizontal, para verificar as dimensões da reprodução do modelo, entre outras (Cunha, 2000a). A pontuação é atribuída em função do sucesso da reprodução, que não está associada à perfeição da realização do desenho, mas se baseia no percentual de acertos observados em diferentes faixas de idade da amostra estudada: *“as notas serão ponderadas em função do valor genético do aspecto ao qual se referem”* (Santucci & Pêcheux, 1981, p.349). As pontuações variam entre um e três. O valor “3” é atribuído aos *“aspectos para os quais os critérios de resultado são raramente respeitados pelas crianças de 6 anos, melhor respeitados de idade em idade ou relativamente pouco respeitados mesmo para crianças mais velhas”* (Cunha, 2000a, p. 302). Já a pontuação “1” é dada *“aos aspectos bem reproduzidos por grande porcentagem de crianças desde os seis anos, ou a certos sucessos de nível inferior, cuja frequência decresce com a idade”* (Cunha, 2000a, p. 302). A pontuação “2” é atribuída aos casos intermediários.

Para a Figura I são avaliados seis itens e a criança pode atingir até 15 pontos; para a II, oito itens e 16 pontos; para a III, sete itens e 12 pontos; para a IV, nove itens e 22 pontos e finalmente para a V, seis itens e 16 pontos. Dessa forma a pontuação máxima possível é de 81 pontos, que, segundo as autoras, não foi obtida por nenhuma das crianças de sua pesquisa mesmo nas faixas de 12 e de 14 anos, em que 30% obtiveram entre 75 e 78 pontos. Os resultados são consultados em uma tabela apresentada em medianas e Quartis para cada idade.

Dessa forma, o sistema proposto por Santucci e Pêcheux (1981) pode oferecer informações importantes sobre a capacidade de organização e o desenvolvimento grafoperceptivo de crianças. As autoras ressaltam ainda a importância de comparar os resultados do teste com provas de nível mental, analisar o escore global do protocolo e observar o comportamento da criança nos contextos de avaliações clínicas.

2.4 – O Teste de Bender como indicador de distúrbios emocionais em crianças: a avaliação Clawson

Clawson (1962/1992) desenvolveu um sistema de aplicação e avaliação do Teste de Bender voltado para a avaliação de crianças, especificamente para aquelas com problemas emocionais, sendo que para sua utilização deve-se primeiro descartar a hipótese de organicidade. Seu foco principal é usar o Bender como um teste projetivo (Cunha, 2000a; Nunes et al.,2007), com a “*intenção de compreender a organização da personalidade da criança*” (Clawson, 1962/1992, p.2).

Para a aplicação, assim como em outros sistemas, deve-se fornecer à criança um lápis com uma borracha na ponta (a autora não recomenda uma borracha separada, pois acredita que seria um estímulo distrator da tarefa) e uma folha de papel para que ela possa iniciar a tarefa (outras folhas podem estar disponíveis, mas não ao alcance). Os cartões devem ser colocados na ordem com o estímulo voltado para cima. São desencorajadas as rotações dos cartões durante a fase de cópia. Caso isso ocorra, é solicitado que a criança desenhe novamente os modelos que foram girados durante a cópia em uma nova folha de papel após a realização do último desenho.

Clawson (1962/1992) propõe a aplicação em quatro fases: 1- Cópia: em que se solicita que os desenhos sejam reproduzidos o mais parecido possível com o modelo; 2- Memória imediata: após a cópia, em uma nova folha, em que se pede à criança que desenhe novamente tantos desenhos quantos recordar; 3- Elaboração: os cartões com os desenhos são apresentados novamente e orienta-se a criança para desenhá-los da forma que quiser, alterando, combinando e/ou elaborando a partir deles e 4- Associações: os cartões são apresentados novamente e pergunta-se à criança com o que eles se parecem ou do que elas lembram ao ver cada desenho.

Clawson (1962/1992) estabeleceu seus critérios de avaliação e interpretação a partir da comparação entre os protocolos de crianças com desenvolvimento normal com outras que apresentavam algum tipo de comprometimento emocional. Sua amostra foi composta por 160 crianças com idades entre 7 e 12 anos, sendo que 80 eram clientes de uma clínica em Wichita (Kansas) consideradas comportamentalmente desajustadas e com perturbações emocionais, mas sem um quadro psicótico, de retardo mental ou de lesão cerebral. O grupo controle foi formado por 80 crianças de escolas públicas, avaliadas por um professor como

normais quanto ao desenvolvimento. Os grupos foram balanceados em função da idade, QI e nível socioeconômico (Cunha, 2000a).

Clawson (1962/1992) destaca a importância de considerar a idade da criança ao interpretar certas alterações na execução da tarefa, uma vez que toda produção infantil pode ser influenciada por fatores do desenvolvimento.

Em relação à fase de associação, Clawson (1962/1992) descreve as respostas mais comuns dadas para cada um dos cartões obtidas nos seus estudos. Também se refere a alterações que normalmente estão associadas a quadros de transtornos cerebrais, destacando, no entanto, que nem todas as crianças com transtorno do sistema nervoso central irão deformar os desenhos, pois em função da grande variedade de transtornos existentes, o teste não diagnostica corretamente todos eles. Entre as características geralmente encontradas destacam-se as simplificações, fragmentações dos desenhos, colisões, rotações, número incorreto de unidades, perseverações, má qualidade de linha, pontos grosseiramente deformados e traços ou vírgulas substituindo pontos.

A autora ressalta a importância da utilização do Bender como uma técnica projetiva, no entanto, destaca a necessidade de seu uso em uma bateria de testes para maior precisão diagnóstica.

2.5 – Adaptação Hutt do Teste Gestáltico Visomotor de Bender para adultos

Hutt (1975) elaborou uma adaptação do Teste de Bender para aplicação em adultos como uma técnica projetiva. Partiu do pressuposto de que os padrões gestálticos percebidos são frutos das experiências pessoais e a forma final da experiência não está somente relacionada à percepção, mas também à personalidade. A avaliação do teste deve considerar tanto o processo de resposta envolvido na reprodução dos desenhos como o produto final, o que possibilitará uma compreensão do comportamento do examinando, do seu *“estilo idiossincrático de personalidade, suas necessidades, conflitos e defesas, seu nível de maturidade emocional, seus métodos de enfrentamento das situações e forças do ego”* (Hutt, 1975, p.12).

Algumas questões merecem destaque na adaptação de Hutt: segundo Nunes et al. (2007), ainda que os desenhos dos cartões sejam os mesmos do original, as linhas e formas foram feitas com maior firmeza e definição (linhas mais finas e

precisas, pontos mais arredondados, etc.). Em sua proposta foram feitas modificações nos procedimentos de administração, porque o objetivo é destacar os aspectos não estruturados da tarefa, facilitando a projeção dos conteúdos inconscientes (Cunha, 2000b; Hutt, 1975, 1985).

A aplicação envolve três fases: cópia, elaboração e associação. Na cópia é solicitado que os desenhos sejam feitos sem o uso de recursos que possam auxiliar a reprodução (uma borracha deve estar disponível para eventuais correções). Na Elaboração, solicita-se nova execução dos desenhos, mas que sejam modificados de acordo com a preferência ou da forma que for mais agradável. Na Associação, os cartões são mostrados novamente junto com a produção feita na elaboração, perguntando-se o que cada desenho lembra tanto em relação ao modelo do cartão como para a elaboração realizada. São propostas outras formas de administração que podem ser acrescentadas: a realização dos desenhos de memória e os testes de limite em que se solicita nova cópia dos modelos que não foram reproduzidos corretamente, que suscitaram dúvidas sobre uma hipótese diagnóstica ou para confirmar uma configuração específica, entre outras.

Hutt (1975, 1985) apresenta duas formas de avaliação: uma qualitativa e outra quantitativa. A qualitativa é voltada para aspectos projetivos / interpretativos do teste, na qual devem ser consideradas todas as fases da aplicação (incluindo a cópia, com base na simbologia implícita em cada desenho). A avaliação quantitativa se refere somente à cópia e é feita com uma escala de aspectos psicopatológicos, elaborada a partir da experiência clínica do autor.

A Escala Revisada de Psicopatologia (Hutt, 1985) é formada por 17 fatores que recebem uma pontuação que pode variar de 1 a 10 em função da severidade de aspectos qualitativos dos resultados (por exemplo, uso do espaço normal ou anormal) e quantitativos (números de ocorrências, por exemplo, mudança na posição da folha uma ou mais vezes, mudanças na angulação em um ou mais desenhos, etc.). Dessa forma, a pontuação pode variar de 17 a 170 pontos em que quanto maior a pontuação maior o grau de psicopatologia. A escala foi elaborada a partir da análise de 775 protocolos de adultos normais e com diferentes quadros clínicos (neuróticos, depressivos, esquizofrênicos e com lesão orgânica).

Além da escala de psicopatologia o autor também propõe uma Análise Configuracional³ (Hutt, 1975, 1985) formada a partir dos itens da Escala de Psicopatologia (com exceção de dois itens: a falta de coordenação da linha e diminuição severa do tamanho de quatro ou mais figuras). Esta análise mostra os itens mais comuns observados em alguns quadros clínicos e apresenta um ponto de corte que permite sugerir com mais segurança a existência destes quadros.

2.6 – Teste Gestáltico Visomotor de Bender - Sistema de Pontuação Gradual (B-SPG)

Sisto et al. (2005) desenvolveram o Teste Gestáltico Visomotor de Bender – Sistema de Pontuação Gradual (B-SPG) partindo dos pressupostos propostos por Bender (1955) de que é possível verificar o nível da maturação da função gestáltica visomotora por meio da reprodução de desenhos. No entanto, enquanto a proposta original de Laretta Bender era mais clínica, eles apresentaram uma avaliação que pode ser aplicada a diferentes contextos e que “*pretende analisar a produção dos sujeitos com maior refinamento e de forma quantitativa*” (Sisto et al., 2005, p. 8). O teste é formado pelos mesmos nove cartões originais, no entanto, são apresentadas novas propostas na forma de aplicação e avaliação do instrumento.

O B-SPG pode ser aplicado tanto individual quanto coletivamente. Na forma coletiva, a aplicação é realizada com o auxílio de um retroprojetor e são utilizadas transparências com os desenhos dos cartões originais que são projetados em uma tela. Nesta modalidade de utilização do teste, o retroprojetor deve ficar a uma distância de no máximo dois metros da tela, os desenhos são apresentados um a um aguardando-se que todos finalizem a cópia para que outro estímulo seja mostrado. É necessária ainda a presença de mais dois auxiliares para acompanhar se a tarefa está sendo feita da forma correta e deve-se acomodar no máximo 30 crianças na sala. No B-SPG as crianças devem copiar os desenhos com a maior precisão que puderem, não podem utilizar borracha, régua ou qualquer outro tipo de objeto como auxílio na execução da tarefa. Apesar dos autores determinarem a distância máxima do projetor à tela nesta modalidade de aplicação, não há referências sobre a interferência da distância em que os examinandos se encontram

³ Tradução não publicada da Profa. Dra. Eda Marconi Custódio, 2003.

da mesma o que pode variar consideravelmente em função da sala onde o teste será administrado.

O objetivo do B-SPG é avaliar a maturidade percepto-motora em crianças com idades entre 6 e 10 anos por meio da análise da distorção da forma dos desenhos, sendo este o único critério adotado na avaliação. Assim, a análise do teste se fundamenta na existência de distorções estruturais dos desenhos apresentados em que os pontos, círculos, linhas, retas, curvas e ângulos são reproduzidos incorretamente (De Filippi & Kato, 2010). Cada desenho recebe uma pontuação que pode variar entre 0 e 3, sendo que quanto maior o valor, maior o número de erros e portanto pior a qualidade da reprodução dos estímulos. Uma pontuação igual a zero significa que o desenho foi feito com uma qualidade excelente, sem erros, enquanto que uma nota três, significa grande quantidade de erros e distorções acentuadas das figuras originais (De Filippi & Kato, 2010). A única figura que pode receber 3 pontos é a 6, as demais podem receber entre 0 e 2 pontos. Quanto à avaliação, na Figura A é avaliado somente o quadrado; as duas figuras geométricas da Figura 7 são pontuadas separadamente e na Figura 8 somente o hexágono é avaliado, desconsiderando o losango central (Sisto et al., 2005). Dessa forma a criança pode atingir um total de até 21 pontos (De Filippi & Kato, 2010).

O manual do B-SPG (Sisto et al., 2005) descreve o estudo realizado para o desenvolvimento do sistema de pontuação proposto. A amostra foi constituída por 1.052 crianças com idades entre 6 e 10 anos com escolaridade desde a pré-escola até a 4ª série do ensino fundamental (atual 5º ano). Também são apresentados os estudos de validade baseados na estrutura interna dos itens obtida por meio do cálculo do Funcionamento Diferencial dos Itens (*DIF*) e na relação com a idade dos participantes (correlações entre a idade e pontuação no B-SPG). O texto apresenta ainda estudos de validade relacionando a pontuação do B-SPG com a escolaridade, comparando grupos extremos com dificuldade de aprendizagem, com os resultados do Teste de Matrizes Progressivas Coloridas de Raven e diferenciação entre grupos extremos, tendo como critério os pontos do Raven. Sobre a precisão do instrumento há estudos de adequação ao modelo Rasch, Consistência Interna (pelo cálculo do Alfa de Cronbach e Spearman-Brown) e Precisão entre Avaliadores.

Os critérios para pontuar as distorções da forma são específicos de cada desenho. Em geral se referem a alterações no tamanho (linhas e ângulos); perda da

forma original substituindo por outra que tenha ou não relação com ela; substituir pontos por círculos e vice-versa; perda de inclinação e/ou direção; omissão ou acréscimo de parte dos desenhos; substituição de ângulos por curvas; simplificação do desenho; alterações na proporção dos elementos do desenho; perda do formato de curvas e perda ou adição de ângulos e linhas.

Sisto et al. (2005) ressaltam ainda que as evidências encontradas nas pesquisas com o B-SPG permitem afirmar que este avalia a maturidade visomotora e está associado a medidas de inteligência como o fator *g*. O B-SPG ofereceria, portanto, “*uma medida de aspectos cognitivos que se relacionam com o fator g e aprendizagem escolar*” (Sisto et al., 2005, p. 98). As normas são apresentadas em função da pontuação bruta e pela medida Rasch. As tabelas foram elaboradas considerando as idades dos participantes e expressam resultados em percentis e quartis. São incluídas ainda tabelas que informam as frequências de pontuações atribuídas por figuras e para o total de pontos do teste em função da idade, além de um estudo de protocolos discrepantes que não se enquadram nos critérios propostos pelo sistema e, portanto, não podem ser interpretados. Os critérios para identificar estes protocolos estão associados à existência de uma discrepância entre a cópia e o desenho esperado em relação à habilidade da criança, de acordo com o modelo *Rasch*.

2.7 – Teste Gestáltico Visomotor de Bender – Segunda Edição (Bender Gestalt II) de Brannigan e Decker

Brannigan e Decker (2003) apresentaram uma nova versão do teste que eles chamaram de Teste Gestáltico Visomotor de Bender – Segunda Edição⁴ (ou Bender Gestalt II). Este material foi desenvolvido tendo como base mais de 60 anos de pesquisas com o teste original. Para sua elaboração foram considerados métodos contemporâneos de construção de testes e padrões atuais de testagem educacional e psicológica. O Bender II foi desenvolvido para ampliar a utilidade do instrumento nas avaliações educacionais, psicológicas e neuropsicológicas (Brannigan, Decker & Madsen, 2004).

⁴ Nesta tese o teste será chamado de Bender II.

O Bender II é uma medida da habilidade de integração visomotora em crianças e adultos com idades entre 4 e 85 anos. Os objetivos da revisão foram: a) aumentar a abrangência da aplicação e da avaliação com a inclusão de itens significativamente mais fáceis e mais difíceis; 2) obter dados sobre as habilidades visuais e motoras em uma amplitude maior de faixas etárias para responder às críticas de que o teste não apresentava normas suficientes para a interpretação dos resultados (para isso foi utilizada uma amostra grande e representativa constituída por 4.000 participantes) e 3) manter o maior número possível de itens do Teste de Bender original.

Em relação ao uso clínico, esta versão apresenta uma escala que avalia o desenvolvimento de defesas e a capacidade de *coping* da personalidade com itens relacionados ao espaço, tamanho, organização, mudanças na Gestalt e distorções. Apresenta também uma Escala de Psicopatologia com itens relacionados a estes três últimos aspectos.

O Bender II é composto por 16 cartões estímulos: foram mantidos os cartões da versão original do teste e acrescentados mais sete, sendo quatro considerados mais fáceis e três mais difíceis com a finalidade de ampliar as possibilidades de avaliação com o instrumento. A sequência dos cartões ficou assim determinada: os quatro novos cartões (considerados mais fáceis) são os primeiros da sequência. O cartão 5 corresponde à Figura A da versão original, seguem-se então os demais cartões originais até o número 13 e posteriormente são apresentados os três cartões considerados mais difíceis (numerados de 14 a 16).

Para a aplicação do teste os cartões são administrados de acordo com a idade do examinando: entre 4 e 7 anos e 11 meses devem ser aplicados os cartões 1 a 13. Já para os examinandos com mais de 8 anos a aplicação deve começar com o cartão 5 e prosseguir até o cartão 16. Dessa forma, no Bender II, todos os cartões originais devem ser aplicados a todos os examinandos independente da idade.

O processo de administração é composto por duas fases: primeiramente solicita-se que o examinando copie em uma folha de papel os desenhos apresentados nos cartões (são mostrados um de cada vez e ficam disponíveis até que a cópia seja finalizada). Os demais cartões ficam sobre a mesa, com a parte da frente voltada para baixo para que não seja visto o desenho que deverá ser feito na sequência. A segunda fase é a da memória, aplicada imediatamente após o término da cópia. Entrega-se uma nova folha ao examinando que deve desenhar novamente

os estímulos sem que eles sejam mostrados, ou seja, devem ser reproduzidos de memória.

O Bender II é formado ainda por duas provas adicionais: um Teste Motor e um Teste de Percepção, utilizados para detectar déficits (ou dificuldades) nas habilidades motoras e perceptuais específicas, que podem afetar adversamente o desempenho da integração visual e motora e interferir nos resultados do teste. Estes testes devem ser administrados após a fase de memória.

O Teste Motor tem a duração de 2 a 4 minutos e a tarefa a ser realizada consiste em desenhar uma linha entre dois pontos em cada uma das situações apresentadas sem tocar as bordas da figura na qual os pontos estão desenhados e sem tirar o lápis do papel. O teste é formado por um exemplo e quatro itens com três figuras cada.

O Teste de Percepção também leva de 2 a 4 minutos. O examinando deve circular (ou marcar) a figura (entre as diversas que aparecem em uma mesma linha) que mais se parece com o modelo. O teste é composto por 10 itens, caso o examinando gaste mais do que 30 segundos para responder a um determinado item, passar ao seguinte.

Além disso, o examinador conta também com um Formulário de Observação para registrar as informações sobre o examinando (sexo, idade, dominância lateral); fatores físicos que podem interferir no desempenho do examinando (deficiências sensoriais, restrições de movimentos); comportamentos durante a realização do teste (descuido, indiferença, falta de atenção, comportamentos específicos e não usuais) e formas/detalhes específicos do ato de desenhar (observações do comportamento e do processo de desenho), bem como para resumir as pontuações das fases de cópia e de memória (tempo necessário para lembrar dos desenhos e ordem em que foram desenhados) e os resultados do teste motor e de percepção.

O Sistema de Pontuação Global usado para avaliar cada desenho feito pelo examinando durante as fases de Cópia e de Memória é uma escala de cinco pontos, sendo que, quanto maior for a pontuação, melhor o desempenho do examinando no teste.

O Sistema de Pontuação Global permite avaliar cada item e a pontuação total do teste, sendo que esta pode variar entre 0 e 52 pontos para examinandos com idade até 8 anos e entre 0 e 48 pontos para os com mais de 8 anos.

O Teste Motor é pontuado pelo seguinte critério:

- Pontuação (1): a linha toca os dois pontos e não ultrapassa os limites da borda (a linha pode tocar a borda, mas não ultrapassá-la);
- Pontuação (0): a linha se estende para fora da borda ou não toca os dois pontos.

A pontuação máxima possível é de 12 pontos.

O Teste de Percepção:

- Cada resposta correta recebe um ponto.
- As respostas incorretas não são pontuadas.

Os examinandos podem atingir no máximo 10 pontos.

Os escores brutos são convertidos em escores padronizados. As normas foram elaboradas em Notas T, Percentis, Intervalos de Confiança e Classificação para as reproduções por faixas etárias. As normas estão disponíveis para as fases de Cópia, de Memória e para os testes Motor e de Percepção (sendo que para os dois últimos são em percentil). Foram elaboradas considerando a idade, o sexo, grupo étnico, a região geográfica e o nível socioeconômico. O manual apresenta estudos com diferentes grupos clínicos (dificuldades de aprendizagem, TDAH, distúrbios emocionais, autismo, Alzheimer, superdotados entre outros).

Brannigan et al. (2004) afirmam que o Bender II pode ser usado nas avaliações de crianças e adultos para os mais diversos fins. Em relação ao grupo infantil é um instrumento útil nas avaliações para a preparação para a escola; para prever o rendimento escolar; diagnosticar problemas de aprendizagem e de leitura e avaliar problemas emocionais e de desenvolvimento. Para adolescentes e adultos o uso é voltado principalmente para a avaliação de lesões cerebrais e diferentes tipos de personalidade; na identificação de distúrbios que podem levar à regressão ou deterioração das gestalts perceptuais e motoras; para diferenciar problemas de aprendizagem, psicológicos e neurológicos.

De acordo com os estes autores, a fase de memória avalia a habilidade do examinando para codificar, armazenar e recuperar as informações, podendo ser influenciada por problemas motores e perceptuais (o que também deve ser considerado para a interpretação da cópia). Nesta parte do teste, também é importante fazer a observação de alguns comportamentos, tais como: o tempo

usado, os níveis de frustração, o número de itens lembrados corretamente e as características inerentes às lembranças. Esta fase tem grande relevância na investigação da doença de Alzheimer, em que frequentemente se encontra um déficit acentuado na sua pontuação, o que não acontece na cópia (Brannigan et al., 2004).

Os resultados quantitativos são baseados no Sistema de Pontuação Global e representam a pontuação do teste. Integram ainda a avaliação, as observações qualitativas de comportamento geral e específico identificados na reprodução dos desenhos.

Os testes suplementares fornecem informações específicas sobre as habilidades motoras e perceptuais, que podem ajudar a explicar baixos resultados nas fases de cópia e de memória. Eles devem ser administrados principalmente quando os escores padronizados estão abaixo do intervalo médio inferior (80-89). Pode-se suspeitar de dificuldades motoras ou de percepção, quando os escores estão abaixo do percentil 25 (Brannigan et al., 2004). Eles destacam ainda outras considerações que são importantes e que devem ser observadas na administração do Bender II:

- a) problemas visuais;
- b) limitações fisiológicas associadas à doenças, lesões, fadiga, fraqueza muscular;
- c) condições fisicamente incapacitantes, tais como o baixo peso ao nascer; paralisia cerebral, ou anemia falciforme;
- d) estresse ambiental;
- e) impulsividade;
- f) motivação inadequada;
- g) problemas emocionais;
- h) deficiência mental;
- i) privação social ou cultural;
- j) experiência limitada.

A proposta de avaliação do Bender II, apresentada por Brannigan e Decker (2003), retoma a importância de uma avaliação qualitativa do teste, além da quantitativa; amplia sua utilização estendendo para uma maior amplitude de faixa etária com o desenvolvimento de novos cartões com desenhos mais simples e mais

complexos, bem como a retomada das ideias de Lauretta Bender (1955/2011) sobre a importância da avaliação do teste todo e não em partes. Este aspecto reforça a ideia de que a função gestáltica inerente à integração visomotora é significativamente mais complexa, do que a percepção visual ou a coordenação motora, tomadas isoladamente (Reynolds, 2007).

2.8 – Sistema de Avaliação do Desenvolvimento de Koppitz para o Teste Gestáltico de Bender – Segunda Edição (Koppitz-2)

Reynolds (2007) realizou uma extensa revisão do Teste Gestáltico de Bender para Crianças de Koppitz (1963/1989) a partir das publicações desta autora e da revisão da versão original do teste proposta por Brannigan e Decker (2003) descrita anteriormente. Elaborou um sistema de pontuação que, como o próprio nome diz, leva em consideração aspectos do desenvolvimento na reprodução dos desenhos e que pode ser utilizado em pessoas com idades entre 5 e 85 anos. A principal diferença entre o Sistema de Pontuação de Desenvolvimento de Koppitz para o Teste Gestáltico de Bender – Segunda Edição⁵ (Koppitz-2) e a avaliação proposta por Brannigan e Decker (2003), se refere às orientações teóricas dos diferentes autores, que resultaram em dois sistemas de pontuação bastante diferentes. A ênfase da avaliação apresentada no Koppitz-2 é a “progressão do desenvolvimento da capacidade de eliminar erros específicos na reprodução dos desenhos” (Reynolds, 2007, p. x).

O sistema original de Koppitz (1963/1989) foi desenvolvido para a avaliação de crianças. Com a revisão realizada para o teste proposta por Brannigan e Decker (2003) na qual foram incluídos novos cartões com desenhos mais complexos, ampliou-se a aplicabilidade do teste em adultos o que acabou motivando o autor a atualizar o sistema de Koppitz para este grupo. Dessa forma pode-se dizer que o Koppitz-2 não é somente uma revisão, mas uma extensão e um redesenvolvimento do sistema de Koppitz original. Reynolds (2007) destaca a importância da avaliação da habilidade de integração visomotora em adultos para diferentes fins, contribuindo para os processos de diagnóstico, especificamente em casos de lesões cerebrais,

⁵ Neste trabalho este sistema será referido como Koppitz-2.

no planejamento de reabilitação profissional e para a crescente população de idosos.

O Koppitz-2 é constituído por 16 cartões com desenhos (os mesmos do Bender II de Brannigan e Decker, 2003), dois formulários de avaliação: um destinado a crianças com idades entre 5 e 7 anos e outro para pessoas com 8 anos ou mais. Também inclui um formulário de indicadores emocionais e um crivo de correção. A aplicação segue a proposta de Brannigan e Decker (2003), ou seja, os cartões de 1 a 13 são aplicados a crianças entre 5 e 7 anos e os cartões de 5 a 16 em pessoas com mais de 8 anos.

Reynolds e Fletcher-Janzen (2000) definem a integração visomotora como a capacidade relacionar estímulos visuais a respostas motoras de forma precisa e apropriada, o que envolve uma grande quantidade de comportamentos visuais e motores existentes em todos os níveis do funcionamento cognitivo. Para Reynolds (2007), na prática, esta capacidade está quase sempre restrita às tarefas que envolvem mais a interpretação ou organização de elementos físicos de um estímulo do que seus aspectos simbólicos e significados. Ele cita um exemplo no qual ao se considerar a capacidade de integração visomotora, é mais importante ficar atento à qualidade ou a forma como uma criança desenha um triângulo que está sendo copiado de um modelo, do que se ela sabe o nome da figura que está sendo reproduzida.

O autor aponta cinco principais usos para o Koppitz-2:

- a) Verificar a existência e o grau de dificuldades visomotoras;
- b) Direcionar encaminhamentos em função dos problemas observados na capacidade de integração visomotora, a ela associados ou dela decorrentes;
- c) Verificar a eficácia de programas de intervenção projetados para corrigir problemas visomotores;
- d) Como instrumento de pesquisa;
- e) Para auxiliar nos diagnósticos diferenciais de várias condições psicológicas e neuropsicológicas relacionados ao avanço da idade.

O teste é aplicado individualmente. O examinador deve separar os cartões que serão aplicados em função da idade. O mesmo deve ser feito para o formulário de registro das informações pessoais e da avaliação. São necessários dois lápis

número 2 com borracha na ponta, cerca de 20 folhas de papel sulfite e um cronômetro, pois apesar do teste não ter um limite de tempo, o tempo de realização deve ser anotado para posterior avaliação. A folha de papel deve ser apresentada na posição vertical, sendo corrigidas todas as tentativas de mudar sua orientação durante a realização do teste. A borracha pode ser usada, assim como um número maior de folhas, no entanto, estes comportamentos não devem ser encorajados (não são mencionados e não fazem parte das instruções de aplicação, mas podem ser autorizados, se solicitado). Assim como nos outros sistemas, os cartões são apresentados e devem ser copiados individualmente sendo registrado o tempo total da aplicação.

A avaliação deve ser feita em formulário específico para a idade, uma vez que de acordo com esta são aplicados desenhos diferentes. No total são avaliados 34 itens nos protocolos das crianças de 5 a 7 anos e 45 itens nos de 8 anos ou mais, sendo estas as pontuações máximas que podem ser atingidas para cada grupo. Cada item recebe o valor "0" se não estiver contemplado, ou seja, se o critério não foi atendido e "1" se a exigência da qualidade da reprodução foi atendida corretamente.

Para os nove cartões originais do teste (desenhos de 5 a 13), que são aplicados nos dois grupos de idade, há algumas diferenças nos critérios de avaliação. Os itens são iguais para os desenhos 6, 9, 11 e 13 em ambos os grupos. Para os desenhos 5 e 8 foi excluído um item no grupo de mais velhos. Nos desenhos 7, 8, 10 e 12 foram incluídos itens mais exigentes na avaliação dos maiores de 8 anos.

Na Escala de Koppitz-2 as pontuações são atribuídas em função da qualidade dos desenhos reproduzidos. Dessa forma, quanto maior a pontuação melhor o resultado no teste. O total de pontos brutos deve ser convertido em percentil e em escore padrão ($M=100$, $DP=15$), sendo este último chamado de Índice Visomotor (VMI). Reynolds (2007) define o VMI como uma pontuação ponderada de desvio da idade corrigida (*age-corrected deviation scaled score*) e deve ser utilizada para a interpretação dos resultados do teste. Para a consulta às tabelas, deve-se calcular a idades do examinando em anos, meses e dias, já que algumas faixas etárias apresentam esta especificação. São apresentadas ainda tabelas para conversão dos resultados em Notas T, escores z, Equivalentes da Curva Normal (*Normal Curve Equivalent* - NCEs) e Estaninos, além de uma tabela de

Equivalentes de Idade (*Age Equivalents*). Apesar de apresentar dados sobre este último tipo de norma (somente entre 5 e 18 anos)⁶, o autor faz ressalvas sobre sua utilização na interpretação dos resultados, afirmando que há muitas divergências e questionamentos sobre sua real utilidade. Dessa forma orienta os examinadores a interpretar os resultados do teste considerando o percentil e o VMI.

A amostra de normatização do Koppitz-2 (Reynolds, 2007) foi a mesma utilizada no Bender II (Brannigan e Decker, 2003)⁷. Constituiu-se de 3.535 participantes com idades entre 5 e 85 anos (ou mais) e foi obtida tendo como base o censo do ano 2000 dos Estados Unidos, procurando controlar as variáveis, sexo, idade, raça/etnia, região geográfica e nível socioeconômico.

Podem ser feitas algumas considerações em relação ao sistema original de Koppitz (1963/1989) e o Koppitz-2 (Reynolds, 2007). Embora em sua avaliação ambos tenham como foco as questões maturacionais, relativas ao desenvolvimento, o foco da Escala de Koppitz é a análise das distorções da forma, as dificuldades de integração, rotação e perseveração, ou seja, os erros cometidos ao reproduzir os desenhos, de forma que quanto maior a pontuação, pior era o resultado, indicando maior imaturidade visomotora. A proposta do sistema Koppitz-2 é pontuar a qualidade da reprodução, ou seja, quanto maior a pontuação, melhor o resultado. Apesar desta diferença, a análise dos dois sistemas mostra semelhança nos critérios de avaliação (embora pontuados de forma inversa), principalmente em relação à distorção da forma e à integração. As alterações relacionadas à rotação e à perseveração bastante destacadas na proposta original da autora, não foram contempladas diretamente nos critérios do Sistema Koppitz-2, podendo apenas ser observadas indiretamente na qualidade da reprodução dos desenhos.

Outra proposta apresentada por Koppitz (1963/1989) foi a avaliação de indicadores emocionais que poderiam aparecer nas reproduções como decorrentes dos problemas visomotores. Na proposta original a autora estabeleceu dez indicadores, que foram mantidos no Sistema Koppitz-2 com o acréscimo de mais dois (caixas em volta do desenho e elaborações ou acréscimos espontâneos), totalizando 12 indicadores. O manual não traz dados normativos sobre os

⁶ A Equivalência de idade foi obtida a partir das médias de pontuação no teste para cada grupo de idade. Somente foram calculadas até os 18 anos, pois os autores afirmam que após esta idade calcular este tipo de dado faz pouco sentido psicológico ou psicométrico.

⁷ Segundo o autor, foi realizada uma parceria entre autores e editoras (Riverside Publishing Company e PRO-ED), que possibilitou que a utilização da mesma amostra do Bender II de Brannigan e Decker (2003) para normatizar/padronizar o Koppitz-2 (Reynolds, 2007).

indicadores emocionais, somente afirma que a presença de mais do que três deles sugere a existência de problemas emocionais significantes. Refere ainda que a interpretação deve ser feita com cautela e associada a outros dados de outras técnicas de investigação.

O manual do Sistema Koppitz-2 (Reynolds, 2007) traz também os estudos de precisão e validade. A precisão foi obtida por três métodos: pela consistência interna para verificar o grau de homogeneidade dos itens; pelo teste-reteste e pela precisão entre avaliadores. O Alfa de Cronbach foi calculado para todas as faixas de etárias do teste e variou entre 0,77 e 0,93. As correlações entre teste-reteste foram obtidas considerando as idades dos participantes e variaram entre 0,75 e 0,84. A precisão entre avaliadores variou entre 0,91 e 0,93. O autor concluiu que os resultados confirmam a precisão do instrumento com coeficientes elevados de acordo com a literatura científica.

O manual apresenta vários tipos de estudos de validade. O primeiro deles é a Curva de Desenvolvimento que mostra a relação entre as pontuações no sistema Koppitz-2 e a idade. O gráfico mostra um aumento da pontuação até 17 anos, uma estabilização nos resultados até os 40 anos e um claro declínio a partir dos 50 anos.

Foram realizadas correlações com os testes WISC-III, WAIS-III e Stanford-Binet V. Em relação ao QI do WISC-III as correlações oscilaram entre 0,34 e 0,63 e, para os índices fatoriais, entre 0,29 e 0,64. Os maiores valores ocorreram com o QI de Execução (0,63) e o Índice de Organização Perceptual (0,64).

Para o WAIS-III as correlações ficaram entre 0,36 e 0,38 para os QIs e 0,29 e 0,36 para os índices fatoriais. Os resultados mais elevados também ocorreram com o QI de Execução (0,38) e com o Índice de Organização Perceptual (0,36).

Em relação ao Stanford-Binet V, para os QIs as correlações variaram entre 0,22 e 0,29, já para os índices fatoriais entre 0,25 e 0,29. De acordo com Reynolds (2007) os resultados mostram que há uma relação entre a integração visomotora e a inteligência, mas que esta não é alta. Esta relação fica mais evidente em testes não verbais em que as tarefas exigem a capacidade visomotora (ou pelo menos ela está associada à atividade a ser realizada para resolver o problema proposto).

Foram realizadas também correlações com o *Woodcock-Johnson III Test of Achievement* (que variaram entre 0,07 e 0,46) e com o *Wechsler Individual Achievement Test – Second Edition* (que oscilaram entre 0,20 e 0,52). Foram conduzidos estudos com o *The Beery-Buktenica Developmental Test of Visual-Motor*

Integration – Fourth Edition (correlações entre 0,46 e 0,51) e entre o sistema original de Koppitz e o Koppitz-2 (-0,88 sem a correção da idade e -0,73 com esta correção).

Para a validade de critério por grupos contrastantes, foram comparadas pessoas com TDAH, dificuldade de aprendizagem (expressa na escrita, leitura e matemática), deficiência mental, autismo e intelectualmente superdotados. Com exceção deste último, em todos os outros grupos a média foi menor do que a do grupo de controle com as mesmas características sociodemográficas. No entanto o autor ressalta que o Koppitz-2 não deve ser utilizado para fins de diagnóstico destes grupos clínicos. A conclusão foi que os resultados obtidos confirmaram a validade do instrumento.

2.9 – Outros Sistemas de Avaliação

Para Piotrowski (1992), o Bender é um instrumento importante nas avaliações psicológicas clínicas, sendo amplamente utilizado principalmente por profissionais da área de saúde mental. No entanto, ele considera que há um uso potencial que não tem sido explorado devidamente: “como um método dinâmico de triagem da personalidade” (p. XIII). Para este autor nos últimos 50 anos, o teste tem sido usado como uma medida perceptivo-motora (principalmente com crianças), um instrumento de avaliação neuropsicológica, da integridade cognitiva e como um teste de personalidade, no entanto, este último uso precisa ser ampliado.

Reichenberg e Raphael (1992) propuseram um sistema de interpretação projetivo dinâmico que proporciona uma avaliação rápida (triagem) da personalidade por meio da avaliação dos nove desenhos individualmente e do protocolo como um todo. Ressaltam, no entanto, que a avaliação exclusiva do teste não deve ser tomada como critério único de diagnóstico ou intervenção clínica.

Baseando-se em sua experiência clínica com a aplicação de mais de 20.000 protocolos de Bender, os autores desenvolveram um Sistema de Interpretação Projetivo de Personalidade, fundamentado em um grupo de teorias diferentes⁸ com o objetivo de oferecer aos clínicos, que desejam aumentar suas habilidades na avaliação das diversas configurações de personalidade, uma nova forma de

⁸ Os autores afirmam que se trata de um sistema pan-teórico, ou seja, baseado em uma compreensão do comportamento humano profundamente enraizado nas obras de muitos teóricos anteriores, evoluindo nos últimos 100 anos de teoria psicodinâmica.

utilização e interpretação do teste. Esta proposta pode ser usada em diferentes contextos tais como: internações médicas e psiquiátricas em hospitais, consultórios particulares, ambulatórios e clínicas de saúde mental, centros de aconselhamento das universidades, programas de avaliação do ensino na rede pública e privada, seleção de pessoal nas empresas e programas de tratamento residencial, entre outros.

Para Reichenberg e Raphael (1992) um tema comum para os pesquisadores contemporâneos da personalidade é a crença de que as produções verbais e escritas, entre outras, refletem o mundo interior ou as percepções inconscientes. A forma como estes processos internos são expressos corresponde ao foco da prática clínica de psiquiatras e psicólogos. O mundo interior de uma pessoa é mais facilmente avaliado por meio das projeções de seus conteúdos, obtidas em situações não estruturadas, como nos testes projetivos.

Schilder (1934, citado por Reichenberg e Raphael, 1992) acreditava na existência de uma relação entre a percepção visomotora e a personalidade. Considerava que a percepção visomotora inclui o funcionamento da personalidade numa tarefa específica e não exclusivamente no ato mecânico da percepção. Por sua natureza não verbal e por ser relativamente pouco influenciável pela cultura, o Bender pode oferecer informações sobre o funcionamento neuropsicológico e da personalidade.

A proposta de Reichenberg e Raphael (1992) é a utilização do Bender como um teste projetivo, ampliando as possibilidades de avaliação da dinâmica da personalidade, aumentando a profundidade e o significado das respostas do teste, uso pouco difundido para o instrumento. Os autores destacam a relevância do teste para identificar conflitos inconscientes em adultos ou em crianças, bem como o entendimento da dinâmica do paciente, aprofundando a capacidade de avaliação da psique humana e contribuindo para um resultado terapêutico positivo. No entanto, ressaltam o cuidado necessário para empregar o teste com esta finalidade, o que exige profissionais treinados. Usado corretamente, o Bender possibilita a obtenção de *“impressões psicodiagnósticas que não seriam conseguidas de outra forma a não ser com meses de trabalho terapêutico investigativo”* (Reichenberg & Raphael, 1992, p. 5).

Neste sistema o teste deve ser aplicado seguindo as orientações de Bender (1955) com algumas alterações: os cartões são colocados em ordem começando

pela Figura A, com a face voltada para cima. São disponibilizadas várias folhas de papel, pode-se permitir que a folha de aplicação seja girada, mas são desencorajadas as rotações dos cartões (o que é interpretado como traço de personalidade manipuladora). Também não é aceito apenas um desenho por folha e não se informa quantos desenhos serão realizados.

Por ser considerado um teste projetivo, procura-se dar poucas informações ao examinando para não interferir no processo de reprodução que será adotado. Nas instruções pede-se que copiem o desenho com atenção e cuidado, procurando reproduzi-lo da melhor forma possível. A tarefa consiste apenas na cópia, podendo ser aplicado a pessoas com idades entre 6 e 92 anos. O tempo de realização dos desenhos, as rasuras e as diferenças das linhas são interpretados, bem como os comentários durante a aplicação. A Tabela 9 apresenta, de forma resumida, a simbologia dos cartões adotada nas interpretações propostas por Reichenberg e Raphael (1992).

De acordo com Reichenberg e Raphael (1992), a primeira publicação de um sistema de pontuação para a avaliação psicodinâmica com o Bender foi a de Pascal e Suttell em 1951. Trata-se de um sistema para adolescentes e adultos, com idades entre 15 e 50 anos que avalia a maturidade para perceber e reproduzir os desenhos sem erros. Segundo Koppitz (1963/1989), Pascal e Suttell *“consideram o desempenho do sujeito neste teste como um reflexo de suas atitudes diante da realidade. Eles veem a capacidade de desenhar as figuras do teste como uma função da capacidade integradora do sujeito ou de sua ‘força do ego”* (p. 17). As distorções são pontuadas de modo que a maior pontuação se refere ao maior desajuste ou perturbação emocional.

Reichenberg e Raphael (1992) elaboraram um Sistema de Interpretação Psicodiagnóstica Avançado para o Bender, com 41 itens referentes a traços de personalidade e do desenvolvimento, baseando-se na interpretação proposta para cada desenho.

Realizaram estudos de validade comparando a sua avaliação do Bender pela escala com a dos terapeutas dos pacientes que preencheram a mesma escala baseando-se no conhecimento que tinham da história de cada um. A fidedignidade do sistema foi pesquisada pela precisão entre avaliadores. Os resultados indicaram adequadas propriedades psicométricas da escala. Os autores também desenvolveram uma lista de erros com 112 itens a serem pontuados e suas

respectivas interpretações. Cada item é associado a um ou mais itens da escala de classificação de 41 itens. A escala é qualitativa (interpretativa), e não um sistema de avaliação referenciado a normas.

Os autores reconhecem as limitações de seu sistema e deixam claro que o objetivo não foi criar um novo teste, mas propor uma avaliação que pode ser útil para profissionais que pretendem avaliar traços de personalidade. Consideram ser um sistema importante para uma triagem, um “mapa inicial”, que pode ser corroborado pelos resultados das medidas de personalidade em uma bateria de testes, principalmente porque a pessoa não tem um controle consciente sobre o teste. Destacam ainda que o fato de ser um sistema projetivo, não o torna menos científico uma vez que *“nenhuma abordagem teórica discerne completamente a complexidade da avaliação da personalidade, os pesquisadores devem estar dispostos a considerar alternativas”* (Reichenberg & Raphael, 1992, p. 117).

3 – Revisão da literatura: pesquisas com o Bender II e o Koppitz-2

Tendo em mente que a proposta do presente trabalho é adaptar o Sistema de Koppitz-2 para uma amostra brasileira foi feito um levantamento bibliográfico por meio do *PsycInfo* com a finalidade de buscar pesquisas com este instrumento na literatura nacional e internacional. Somente foram encontradas três referências, sendo um artigo que faz uma descrição do teste (Gorske, 2007), outro que se refere a um estudo de validade realizado com crianças com Transtorno do Espectro do Autismo de Alto Funcionamento (McDonald et al., 2014) e um resumo no *Dissertation Abstracts Internacional* sobre o Koppitz-2 como um instrumento de triagem entre crianças e adolescentes (Marinc, 2011).

Uma vez que o sistema de Koppitz-2 (Reynolds, 2007) representa outra forma de avaliação dos 16 cartões propostos por Brannigan e Decker (2003), como citado anteriormente, optou-se por fazer um levantamento bibliográfico dos estudos destas duas versões. Os trabalhos publicados empregaram principalmente amostras de crianças e se referem à validade do instrumento, pela correlação com outros instrumentos que avaliam a mesma característica, construtos relacionados ou diferentes grupos clínicos (Allen & Decker, 2008; Brannigan & Brannigan, 1995; Brannigan & Brunner, 1991; Brannigan & Brunner, 1993; Moose & Brannigan, 1997 e Volker et al., 2010).

Brannigan e Brunner (1991) pesquisaram a validade de construto do Sistema de Pontuação Qualitativa da Versão Modificada do Teste Gestáltico de Bender por meio da correlação com a Escala Maturacional de Koppitz (1963/1974). Essa versão proposta por Brannigan e Brunner (1989, 2002), é um refinamento e ampliação da pesquisa de Jansky e de Hirsch (1972, citados por Brannigan & Brunner, 2002) motivada pela necessidade de avaliar a efetividade do Bender junto com outras medidas de avaliação educacional na predição do desempenho escolar principalmente de crianças pequenas. Nesta versão que deve ser utilizada apenas com crianças entre 4 anos e 8 anos e meio, somente seis cartões são aplicados: Figura A, 1, 2, 4, 6 e 8. O sistema de pontuação é uma escala que varia de “0” (zero) a “5” (cinco) pontos. A pontuação “0” é atribuída quando o desenho é aleatório, rabiscos, com a perda do conceito do desenho e a pontuação “5” é atribuída quando é realizada uma representação acurada do desenho. Dessa forma, a pontuação máxima obtida neste sistema de avaliação qualitativa é de 30 pontos.

Os participantes foram 74 crianças do jardim da infância, 94 da primeira série e 92 da segunda. Todas foram submetidas ao Bender (somente foram aplicados os seis desenhos desta versão do teste) e os protocolos foram avaliados pelos dois sistemas. Para o Sistema Qualitativo foi atribuída uma pontuação de até cinco pontos para cada desenho. Já para a Escala Maturacional foram pontuados os erros de distorção, integração, rotação e perseveração na reprodução dos desenhos. Os resultados mostraram correlações moderadas e altas (Dancey & Reidy, 2006, 2013; Sisto, 2007) e estatisticamente significantes ($p < 0,001$) entre os dois sistemas: -0,80, -0,78 e -0,67 respectivamente para o jardim da infância, a primeira e a segunda séries (as correlações são negativas, pois no sistema qualitativo quanto maior a pontuação melhor a qualidade da reprodução e na escala maturacional a pontuação é atribuída aos erros, assim quanto maior a pontuação pior o resultado). Os autores concluem que os resultados apoiam a validade do Sistema Qualitativo como uma medida da integração visomotora.

Brannigan e Brunner (1993) compararam o Sistema de Pontuação Qualitativo da Versão Modificada do Bender e a Escala Maturacional de Koppitz em estudantes de primeira e segunda séries, com uma dupla finalidade: verificar a relação entre os dois sistemas e identificar se os mesmos podem ser preditivos do desempenho acadêmico, bem como verificar qual dos dois sistemas mostraria uma melhor relação com os resultados de um teste de aptidão escolar, o Teste de Aptidão Escolar de

Otis-Lennon (*Otis-Lennon School Ability Test*). Justificam este estudo com base em pesquisas que mostram que a habilidade de integração visomotora parece contribuir para o desempenho acadêmico. Apoiam-se também em uma afirmação de Bender de que o desempenho no teste “*envolve algumas funções fundamentais da inteligência que são importantes para as habilidades de aprendizado da criança na escola*” (Bender 1938, citado por Brannigan & Brunner, 1993, p.327). Foi usada a Versão Modificada do Bender (Brannigan & Brunner, 1989), que avalia a maturação desta habilidade em crianças entre as idades de 4 anos e 6 meses e 8 anos e 5 meses, pois um desempenho pobre no teste pode refletir atrasos no desenvolvimento e/ou dificuldades de aprendizagem. Embora alguns estudos mostrem que os dois sistemas se correlacionam significativamente com testes de rendimento escolar, os resultados têm sido melhores com o Sistema de Pontuação Qualitativo.

A amostra foi formada por 75 estudantes da primeira e 84 da segunda série de escolas do ensino fundamental de Nova York, que foram submetidos à Versão Modificada do Bender e ao Teste de Aptidão Escolar Otis-Lennon. Os protocolos foram avaliados pela Escala Maturacional de Koppitz (1963/1974) e pelo Sistema de Pontuação Qualitativa (Brannigan & Brunner, 1989). O Teste de Aptidão Escolar Otis-Lennon é uma medida das capacidades importantes para o desempenho escolar: facilidade das crianças para fazer analogias, classificar, seguir instruções, raciocinar com elementos quantitativos e compreender conceitos verbais. O resultado do teste é um escore composto chamado de índice de aptidão escolar. Para verificar os objetivos propostos, foram calculadas as correlações dos dois sistemas de avaliação com o teste de aptidão escolar e, posteriormente, empregado um teste de diferenças entre correlações dependentes para averiguar, se a diferença entre os escores dos dois sistemas era significativa. Os resultados mostraram que para os participantes da primeira série, as correlações foram moderadas e significantes ($p < 0,001$), sendo 0,55 para o Sistema Qualitativo e -0,52 para a Escala Maturacional. Para a segunda série, as correlações foram menores, 0,38 para o Sistema Qualitativo ($p < 0,01$) e -0,20 ($p < 0,05$) para a Escala Maturacional. As correlações com o Sistema Qualitativo foram significativamente maiores do que com a Escala Maturacional para os estudantes da segunda série, mas não para a primeira. As diferenças entre as correlações foram estatisticamente significantes indicando que o Sistema Qualitativo parece ser “um recurso valioso para a

identificação de crianças com problemas escolares” (Brannigan & Brunner, 1993, p. 329). Concluem que os resultados reforçam as críticas à Escala Maturacional e a outros sistemas de avaliação, porque estes acabam separando os fatores perceptivos, motores e maturacionais envolvidos na reprodução dos desenhos, perdendo a noção do todo (da *gestalt*) que é fundamental na proposta do teste. O diferencial do Sistema Qualitativo é a avaliação global que tem maior relação com os fundamentos da Gestalt e com o grau de diferenciação dos desenhos.

Brannigan e Brannigan (1995) compararam a aplicação individual e em grupo da Versão Modificada do Bender. A amostra foi constituída por 90 crianças com escolaridade entre o jardim da infância e a segunda série de diferentes escolas, procurando manter as mesmas características dos grupos no que diz respeito ao sexo, idade e nível socioeconômico dos participantes. Em uma escola o teste foi aplicado individualmente e na outra com slides. A comparação das médias dos resultados dos dois grupos não mostrou diferenças estatisticamente significantes ($t=0,04$; $p>0,05$) entre a aplicação individual ($M=20,11$; $DP=2,81$) e grupal ($M=20,08$; $DP=2,08$).

Moose e Brannigan (1997) compararam os resultados de 50 crianças pré-escolares, sendo 25 de cada sexo, com idade média de 4 anos e 10 meses na Versão Modificada do Bender e o Teste de Desenvolvimento da Integração Visomotora – VMI (Beery, 1989). As crianças foram testadas como parte da avaliação pré-escolar. A ordem da aplicação foi contrabalanceada de forma que metade da amostra fez primeiro a Versão Modificada do Bender e depois o Teste de Integração Visomotora e na outra metade a ordem foi invertida. Os resultados mostraram que a média para a Versão Modificada (em função da Nota T) foi de 45,5 ($DP = 10,08$) e para o Teste de Desenvolvimento da Integração Visomotora - VMI (em função dos escores padronizados) foi de 99,1 ($DP = 11,4$). A correlação entre os dois testes foi de 0,54 ($p<0,01$). Os autores concluíram que o resultado indica a validade para a Versão Modificada do Bender.

Allen e Decker (2008) investigaram o uso do Bender II na avaliação do Transtorno de Déficit de Atenção e Hiperatividade – TDAH, especificamente em relação à impulsividade e aos problemas de atenção, comportamentos comuns neste grupo clínico. Constataram que a possibilidade de avaliar a impulsividade e os problemas de atenção pelo Bender, mesmo em diferentes sistemas de avaliação, não é conclusiva. Em geral os estudos apresentam problemas metodológicos que

dificultam a generalização dos resultados. Citam ainda a falta de uma fundamentação teórica clara dos dois construtos investigados e, especificamente em relação ao uso do Bender, o fato do teste (em seu formato original) ser formado por estímulos bastante simples, fazendo com que eles possam ser facilmente reproduzidos com alta qualidade por crianças entre 10 e 11 anos (que apresentam um desempenho muito semelhante ao de adultos), e somente em casos mais graves, podem ser detectados problemas nas reproduções dos desenhos. Procuraram verificar como a impulsividade e os problemas de atenção em portadores de TDAH surgem nos resultados do Bender, isto é, se as características dos desenhos que são interpretadas como traços de impulsividade e falta de atenção aparecem nos protocolos do Bender de crianças com TDAH.

Como o conceito de TDAH também é bastante discutido na literatura científica, eles adotaram o modelo proposto por Barkley (1999, citado por Allen & Decker, 2008) que tem como foco o estudo da inibição do comportamento, das funções executivas e da autorregulação. De acordo com o modelo de Barkley, a inibição do comportamento possibilita o funcionamento de quatro funções executivas específicas: a memória de trabalho verbal, a não verbal, a autorregulação e a reconstrução. O intervalo de tempo para a resposta no caso de inibição do comportamento direciona o indivíduo para um objetivo melhor do que uma resposta impulsiva. Neste intervalo de tempo, a memória de trabalho verbal (fala internalizada) ajuda a moderar o processamento das outras funções executivas, o que aumenta as chances de obter um comportamento adaptado, fundamentado com um propósito e orientado para o futuro. Nesta proposta, a inibição do comportamento e um eficiente uso das funções executivas, levam a um melhor controle motor. Como na literatura são encontrados vários estudos que relacionam o TDAH com o déficit motor, este fato justificou a sua pesquisa, pois estes déficits devem afetar a desempenho no Bender, uma vez que o teste exige “*habilidades de memória de trabalho intactas, habilidades de planejamento e análise da reconstituição de componentes, a autorregulação que interfere na forma como os desenhos estão sendo copiados e, finalmente, o controle motor*” (Allen & Decker, 2008, p. 665).

O objetivo do estudo foi investigar três hipóteses principais: 1- Crianças com TDAH teriam um desempenho pior do que o grupo de controle nas fases de cópia e de memória do Bender II, quando controlado o efeito do QI (avaliado pela Escala de Inteligência Stanford-Binet – 5ª Edição); 2- Os indicadores de impulsividade, falta de

atenção e planejamento pobre de Koppitz (1963, 1975) e Hutt (1969, 1985), ocorrem com maior frequência em sujeitos com TDAH, quando comparados aos que não têm esta queixa e 3- Os protocolos do Bender II das crianças com TDAH apresentam mais indicadores de TDAH do que o grupo controle.

A pesquisa foi realizada com 124 participantes, sendo 62 com diagnóstico de TDAH (45 do sexo masculino e 17 do feminino com idades entre 5 anos e 8 meses e 18 anos e 2 meses) e 62 sem esta queixa (33 do sexo masculino e 29 do feminino, com idades entre 5 anos e 11 meses e 18 anos e 10 meses). No grupo clínico, 36 participantes foram diagnosticados com TDAH do tipo combinado (falta de atenção e hiperatividade-impulsividade); 16 do tipo predominantemente desatento e nove do tipo predominantemente hiperativo e impulsivo. Os Indicadores Emocionais de Koppitz pesquisados foram: ordem confusa, substituição de traços por círculos, reforçamento de linhas, segunda tentativa para realizar o desenho e expansão / uso adicional do papel. Os Indicadores Emocionais de Hutt foram: colisão, simplificação e a elaboração/rabiscos. Os indicadores de TDAH observados foram: os déficits na memória de trabalho, na capacidade de reconstituição (análise e síntese), na autorregulação / automonitoramento e na capacidade de planejamento, além de um prejuízo global na fase de memória.

Os dois grupos se diferenciaram de forma estatisticamente significativa ($p=0,03$) em relação ao QI avaliado no Stanford-Binet (a média dos participantes com TDAH foi menor) o que também foi observado na pontuação do Bender II ($p<0,001$), tanto na fase de cópia (a média do escore padronizado do grupo com TDAH foi de 95,7 (DP=12,3) e o grupo sem TDAH 103,8 (DP=12,6), quanto na de memória (com TDAH 94,6 (DP=13,6) e sem TDAH 101,6 (DP=15,3). Como neste sistema as pontuações são maiores quanto melhor for a qualidade dos desenhos, pode-se verificar que os resultados piores foram dos participantes com TDAH. Os autores também encontraram correlações moderadas e significantes ($p<0,001$) entre o QI e o Bender II (respectivamente 0,41 e 0,50 para as fases de cópia e de memória).

Para verificar a primeira hipótese, as análises foram repetidas controlando o efeito do QI dos participantes (como covariável), as diferenças entre os grupos na ANCOVA foram significantes para a pontuação na fase de cópia, no entanto, o

tamanho do efeito foi pequeno⁹. Para a fase de memória os resultados não foram significantes, nem para o total de pontos, nem para o número de desenhos realizados (para estes cálculos foram usados somente os protocolos dos participantes entre 8 e 18 anos, pois o número de desenhos aplicados é diferente conforme a idade) para os participantes com e sem TDAH. Para a segunda hipótese, dos oito indicadores de Koppitz e Hutt para impulsividade, desatenção e planejamento somente um deles diferenciou significativamente os dois grupos: a ordem confusa foi mais frequente no grupo com TDAH. Para a terceira hipótese, os indicadores de TDAH utilizados foram: comportamentos de planejamento observados pelo examinador, o uso de linhas auxiliares, apagar ou não o desenho e o tempo para a conclusão da tarefa. Somente foram avaliados os protocolos dos participantes entre 8 e 18 anos. Não foram encontradas diferenças significantes entre os dois grupos em relação aos indicadores de TDAH.

Os autores concluem que, ainda que tenham sido observadas diferenças entre as médias dos grupos na fase de cópia no que diz respeito ao controle motor, capacidade de planejamento e autorregulação, no resultado global, estas diferenças diminuem com o controle do QI (o tamanho do efeito é muito pequeno, ainda que as diferenças sejam significantes). O mesmo ocorreu em relação à fase de memória, em que era esperado um resultado pior do grupo com TDAH em relação à reprodução dos desenhos e ao número de desenhos lembrados. Ainda que tenham sido observadas diferenças no sentido esperado, conforme a literatura científica, ao se controlar o efeito do QI, as diferenças passaram a não ser significantes. Quanto aos indicadores de Hutt e Koppitz, que teoricamente estão associados à impulsividade e ao déficit de atenção, somente a ordem confusa diferenciou os grupos, tendo uma relação estatisticamente significativa com o TDAH. Este resultado sugere a sensibilidade do Bender II para identificar déficits de planejamento e reconstrução nos portadores de TDAH. Os indicadores de TDAH também não diferenciaram os grupos estudados. Apesar dos resultados não confirmarem a utilidade do Bender II na avaliação do TDAH, os autores apontam para limitações do estudo, como o número reduzido de participantes e a necessidade de outras pesquisas considerando os diferentes grupos de TDAH separadamente.

⁹ O tamanho do efeito se refere à porcentagem da variância que pode ser explicada pela variação da variável dependente. No caso apenas 7% da variância dos escores do Bender II podem ser atribuídos a pertencer ou não ao grupo de TDAH.

Allen e Decker (2008) afirmam ainda que os pesquisadores não têm um consenso sobre as técnicas que podem auxiliar a avaliar o TDAH e fazem críticas ao Bender dizendo que o instrumento tem uma história questionável relativa à utilização em desordens psicológicas e para o TDAH. Questionam ainda a eficácia dos indicadores emocionais de Koppitz por apresentarem pouca fundamentação teórica sobre sua relação com o comportamento.

Volker et al. (2010) compararam o Bender II e o Teste do Desenvolvimento da Integração Visual-Motora Beery-Buktenica (VMI-V) em uma amostra de crianças normais e outra com Transtorno do Espectro do Autismo de Alto Funcionamento. Segundo os autores, a escolha dos instrumentos foi baseada na frequência de seu uso nas avaliações psicoeducacionais das habilidades motoras finas e visoperceptivas. Embora a tarefa nos dois testes seja copiar figuras geométricas e avaliar a habilidade visomotora, existem diferenças no conteúdo, na administração e na pontuação dos instrumentos. Por isso verificaram tanto a diferença entre os dois grupos nos testes para a discriminação dos participantes autistas em relação ao grupo controle quanto os resultados do mesmo grupo nos dois testes. O grupo clínico se enquadra em uma categoria mais ampla de Transtornos Invasivos do Desenvolvimento. Ainda que não apresentassem alguns transtornos típicos do autismo (como prejuízos no funcionamento social e comunicativo e comportamentos restritos, repetitivos e estereotipados), as deficiências motoras e a falta de destreza são frequentemente observadas nesse grupo. O objetivo foi ampliar as pesquisas para conhecer mais detalhadamente os déficits motores associados a este quadro clínico. Os participantes foram selecionados de forma a constituir um grupo com Transtorno do Espectro do Autismo de Alto Funcionamento (grupo clínico), um grupo controle e um terceiro formado por participantes dos dois outros grupos.

O grupo clínico foi constituído por 60 crianças com idades entre 6 e 14 anos ($M=9,7$; $DP=2,1$), a maioria do sexo masculino, com QIs entre 71 e 131 ($M=98,5$ e $DP=14,3$), obtidos por meio do WISC-IV. Os critérios de inclusão na amostra foram: diagnóstico de transtorno do espectro autista, ausência de déficit cognitivo substancial e de atraso da linguagem. Os participantes deste grupo passaram por um processo bastante rigoroso e preciso de avaliação para controlar as variáveis que eventualmente poderiam interferir na pesquisa. O grupo de controle foi formado por 46 participantes com idades também entre 6 e 14 anos ($M=9,8$; $DP=2,1$), metade de cada sexo, com QI entre 90 e 132 ($M=109,0$; $DP=10,0$). Neste grupo, os

participantes eram provenientes de escolas de ensino regular e sem histórico de diagnóstico psiquiátrico ou envolvimento em educação especial. O terceiro grupo foi formado por 54 participantes selecionados dos grupos clínico e de controle originais (27 de cada) tendo o cuidado de que apresentassem a mesmas proporções em função do sexo, raça e escolaridade dos pais. Foram mantidas ainda as distribuições das idades e somente o QI obtido pelo WISC-IV foi significativamente maior para o grupo controle.

O VMI-V é um teste para identificar dificuldades na coordenação de habilidades motoras e perceptivo-visuais em crianças e adolescentes com idades entre dois e 18 anos. É composto por 30 itens, sendo 24 figuras geométricas organizadas em ordem crescente de dificuldade. Os desenhos devem ser copiados a lápis, em locais específicos e sem usar a borracha. A avaliação é obtida pela pontuação individual nos desenhos, atribuída de acordo com os critérios pré-definidos pelos autores (Beery & Beery, 2004). O VMI-V apresenta ainda dois testes complementares que podem explicar um mau desempenho no teste: um Teste de Percepção Visual de três minutos em que deve ser identificada a correspondência exata para até 27 figuras geométricas e o Teste de Coordenação Motora com um limite de 5 minutos para traçar até 27 estímulos sem sair do espaço que é delimitado por duas linhas em cada um deles. Os resultados brutos são convertidos em escores padronizados e corrigidos em função da idade.

O Bender II foi aplicado e avaliado de acordo com a proposta de Brannigan e Decker (2003). Os participantes realizaram as fases de cópia, memória e as duas provas adicionais: o Teste Motor e de Percepção. O VMI-V foi administrado seguindo as orientações do manual do teste (Beery & Beery, 2004). A ordem de aplicação foi estabelecida de forma que metade da amostra realizou primeiro o Bender II e a outra metade primeiro o VMI-V. As avaliações, foram realizadas por dois profissionais independentes e as eventuais discrepâncias foram solucionadas por um terceiro avaliador (a precisão entre diferentes avaliadores, atingiu escores de concordância acima de 0,92).

Volker et al. (2010) utilizaram três procedimentos de análise: 1) Comparar os resultados de cada grupo (clínico e controle) no Bender II e no VMI-V; 2) Para o terceiro grupo foram comparados os resultados nos dois testes (incluindo as tarefas complementares/adicionais) para verificar se os mesmos discriminam os

participantes dos grupos clínico e de controle e 3) Correlações entre todos os escores visomotores de cada um dos grupos clínico e controle.

Os resultados mostraram que nos dois grupos na comparação entre a fase de cópia dos dois testes, a média do Bender II foi maior. Entre os testes complementares do VMI-V, a média da Percepção Visual foi maior do que a da Coordenação Motora. Em relação aos resultados do Bender II, a média da fase de cópia foi maior do que a de memória (os testes complementares do Bender II não foram comparados por causa do efeito de teto e a falta de um escore padrão que possibilitasse essa comparação). Em todas as comparações as diferenças foram estatisticamente significantes ($p < 0,01$). As médias do grupo clínico foram menores do que as do controle. Para assegurar maior precisão dos resultados, foram realizados dois tipos de cálculos: considerando ou não o efeito do QI dos participantes (uma vez que a diferença entre essas médias para os grupos de controle e clínico foi estatisticamente significativa). Quando não se considera o controle do QI não foram observadas diferenças estatisticamente significantes entre as médias nos testes de Percepção, na fase de memória do Bender II e no Teste de Percepção Visual do VMI-V. Os grupos se diferenciaram estatisticamente na fase de cópia no Teste Motor do Bender II e na fase de cópia e no Teste de Coordenação Motora do VMI-V. As médias foram maiores para o grupo controle com tamanho de efeito alto nos resultados. Os resultados foram semelhantes nas análises pela ANCOVA, controlando o efeito do QI.

Em relação às correlações entre os dois testes, para o grupo de controle elas variaram entre 0,17 e 0,50 e para o clínico entre 0,25 e 0,56, evidenciando coeficientes de validade adequados para os dois testes nas duas amostras. As maiores correlações ocorreram entre a fase de cópia do Bender II e o Teste de Coordenação Motora do VMI-V (0,56 para o grupo clínico e 0,50 para o controle) e entre as fases de cópia do Bender II e do VMI-V (0,55 para o grupo clínico e 0,48 para o controle). Em relação à fase de memória do Bender II, a correlação com a fase de cópia do mesmo teste foi de 0,49 para o grupo clínico e 0,48 para o controle e com a cópia do VMI-V 0,46 para o controle e 0,34 para o clínico.

Os autores destacaram também que, uma vez que os resultados mostraram que as médias dos escores padronizados do Bender II e do VMI-V são consideravelmente diferentes (a média do primeiro é superior à do segundo), isso sugere que os dois instrumentos tendem a oferecer escores absolutos

consideravelmente diferentes. Por este motivo eles alertam para a importância da escolha do teste a ser usado nas avaliações, pois as conclusões gerais sobre as habilidades visuais e motoras podem ser influenciadas pelo teste empregado.

Em um estudo para verificar a relação entre a habilidade visomotora e uma tarefa de discriminação perceptual que exige a capacidade de atenção sustentada, Soto (2011) aplicou em um grupo de 90 crianças com idades entre 5 e 8 anos o Teste de Bender Modificado e o Teste de Percepção de Diferenças, Rostos que é um teste que avalia a percepção rápida de semelhanças e diferenças em rostos (olhos, boca, cabelos, etc). Os testes foram aplicados coletivamente (foi usado um caderno com as imagens do Bender que era fornecido para a cópia) e avaliados para a obtenção do total de pontos. As análises estatísticas apontaram uma ausência de relação entre a atenção medida no Teste de Rostos e o desempenho no Bender. Dessa forma, os problemas na motricidade e percepção visual identificados no Bender podem não contribuir para compreender o funcionamento atencional de crianças. Concluiu que estes resultados contribuem para a ideia de que o Bender teria uma relação pequena com tarefas que exigem velocidade de processamento. Ainda que os componentes visuais e motores do Bender apresentem uma relação com o desempenho cognitivo em algumas tarefas, a alta especificidade de atividades relacionadas à atenção sustentada e velocidade de processamento parece interferir nesta relação.

Marnic (2011) investigou a possibilidade de utilizar o Sistema de Pontuação Global do Bender II (Brannigan e Decker, 2003) e o Koppitz-2 como um instrumento de triagem na tomada de decisões em diagnósticos psicológicos. Os resultados das avaliações destes dois sistemas foram correlacionados com *Kiddie Schedule for Affective Disorders and Schizophrenia for School Aged Children, Present and Lifetime Edition (KSADS-PL+)*, uma entrevista semiestruturada junto com o *Longitudinal Evaluation of all Available Data (LEAD)* e o *Child Behavior Checklist (CBCL)*. A amostra inicial foi constituída de 115 crianças e adolescentes, no entanto somente 75 completaram todos os instrumentos, sendo estes os considerados para a pesquisa. Os resultados foram correlacionados e foi realizada uma análise da *Curva ROC*, que é um método estatístico utilizado em testes diagnósticos, pois fornece informações sobre sua sensibilidade e especificidade (verdadeiros/falsos positivos e verdadeiros/falsos negativos). Os resultados dos dois sistemas mostraram correlações moderadas com o *CBCL* nas categorias dos sintomas de

comportamentos agressivos e depressivos e Transtorno de Déficit de Atenção e Hiperatividade (TDAH) em crianças e adolescentes (nenhuma das correlações com os outros instrumentos foram estatisticamente significantes). No entanto, os Indicadores Emocionais do Koppitz 2 não discriminaram a presença ou ausência de psicopatologias. As análises da Curva *ROC* de sensibilidade e especificidade mostraram que o uso dos indicadores emocionais do Koppitz-2 melhoraram a taxa de probabilidade de diagnóstico do Transtorno de Déficit de Atenção e Hiperatividade (TDAH) entre 50% e 66%. A pesquisadora concluiu que o uso do Koppitz-2 no diagnóstico do TDAH indica uma possível direção para pesquisas futuras.

Soto (2013) estudou a dimensionalidade do Teste Gestáltico Visomotor de Bender Modificado por meio da Modelagem de Equações Estruturais. O autor justifica a importância da pesquisa pelo fato de que um levantamento sobre os estudos de validade com o teste mostrou que os que se referem a dimensionalidade da pontuação são os menos frequentes, sendo mais encontrados os que têm como objetivo verificar a predição do instrumento e a correlação/convergência com outras medidas visomotoras. A amostra foi formada por 1.381 crianças da cidade de Lima no Peru, de ambos os sexos, com idades entre 4 e 9 anos, que fizeram a versão modificada do Bender (Brannigan & Brunner, 2002) em que são aplicados apenas seis dos nove desenhos originais (A, 1, 2, 4, 6, e 8). A avaliação consiste em atribuir uma pontuação que varia de 0 a 5 conforme a qualidade da reprodução feita pelo examinando. Os testes foram aplicados individual e coletivamente e avaliados por profissionais treinados no sistema de avaliação do Bender Modificado. Foram conduzidas diferentes análises estatísticas (análises fatoriais exploratórias, confirmatórias, cálculos de consistência interna entre outros). O autor concluiu que os resultados apontaram a validade da estrutura interna da versão modificada do Bender, confirmando sua utilidade como um instrumento para a avaliação da habilidade visomotora e que foram encontradas evidências da possibilidade de usar o teste para avaliar pessoas mais velhas.

McDonald et. al (2014) investigaram a relação entre o Bender II com o sistema Koppitz-2 e o *Beery-Buktenica Developmental Test of Visual Motor Integration-6th Edition* (VMI-VI) com o intuito de ampliar as pesquisas sobre as medidas visomotoras e as características do Transtorno do Espectro do Autismo de Alto Funcionamento (HFASD). Este transtorno é caracterizado por dificuldades para

estabelecer relações sociais e interesses e comportamentos restritos e repetitivos. Apesar de nem sempre apresentarem dificuldades nas capacidades cognitivas e de linguagem, jovens com HFASD, frequentemente apresentam outros tipos de problemas que incluem déficit de motricidade fina e grossa e prejuízos nas habilidades visomotoras (McDonald et. al, 2014). Os participantes foram avaliados por estes dois instrumentos e pelos testes complementares do Bender II (Teste Motor e de Percepção de Brannigan & Decker, 2003). Para atingir o seu objetivo compararam os resultados de crianças e adolescentes com e sem HFASD. O grupo foi formado por 141 crianças, sendo 90 com HFASD e 51 sem este transtorno para as comparações intragrupo. As idades dos participantes com HFASD variaram entre 6 e 16 anos ($M=9,41$), sendo a maior parte do sexo masculino (91%). O grupo sem HFADS foi formado por crianças e adolescentes entre 6 e 15 anos ($M=9,88$), sendo 51% do sexo masculino. Uma subamostra de 33 participantes foi composta por pares com e sem HFASD em que foi controlada a idade, sexo, etnia, educação dos pais e a capacidade verbal, com o objetivo de controlar variáveis que eventualmente pudessem interferir nos resultados dos testes. Os dados deste grupo foram utilizados para determinar a diferença entre grupos em relação ao potencial visomotor. Os resultados mostraram que nas comparações intragrupos (independente do grupo) os participantes tiveram um resultado significativamente menor no VMI-VI do que no Koppitz-2, o que também ocorreu em relação ao teste complementar de Coordenação do VMI-VI, comparado ao de Percepção. Além disso, as correlações entre o VMI-VI, os testes complementares do VMI-VI (Coordenação e Percepção) e o Koppiz-2 variaram entre 0,44 e 0,56 ($p<0,01$) para o grupo com HFASD e entre 0,36 e 0,50 para os participantes sem este quadro o que, segundo os autores, representa uma evidência de validade convergente moderada entre os instrumentos. Para os resultados das subamostras, as comparações entre os grupos mostraram diferentes padrões no desempenho nas tarefas que exigiam as habilidades visuais e motoras. Em geral, não foram encontradas diferenças estatisticamente significantes entre as médias dos grupos nos testes em que as exigências motoras eram menores (Percepção Visual do VMI-VI e o Teste de Percepção do Bender II), já no que diz respeito aos testes em que a influência motora era maior (Coordenação Motora do VMI-VI, Koppitz-2 e Teste Motor do Bender II) os resultados foram significantes com tamanhos de efeitos que oscilaram entre moderado e grande. Em todas as comparações, as médias dos participantes

com HFASD foram menores do que a dos outros sem este quadro. Os autores afirmaram que os resultados foram semelhantes aos obtidos por Volker (2010) entre o Sistema de Pontuação Global do Bender II e o VMI-VI. Outra observação realizada por eles é a de que quanto maior for a exigência motora na realização de alguma atividade ou tarefa, há uma tendência ao menor desempenho dos jovens com HFASD.

4 – Pesquisas sobre a relevância do uso do Bender como instrumento de avaliação psicológica

Em um estudo que teve como objetivo identificar os instrumentos psicológicos mais conhecidos e mais utilizados por estudantes e profissionais de Psicologia, Noronha, Beraldo e Oliveira (2003), verificaram que o Teste de Bender foi um dos 15 mais citados (entre os estudantes ocupou o quarto lugar sendo mencionado por 91,46% dos participantes e para os profissionais em sétimo lugar, apontado por 88,5%). Noronha, Primi e Alchieri (2005) realizaram um estudo semelhante com estudantes e psicólogos de diferentes estados do Brasil e o Bender também foi citado entre os instrumentos mais apontados pelos participantes (principalmente em São Paulo, Mato Grosso do Sul e Rio Grande do Sul). O Bender também foi um dos testes mais ensinados em um estudo realizado por Alves, Alchieri e Marques (2002) que abrangeu professores de 62 cursos de graduação de Psicologia.

Atualmente, somente um sistema de avaliação do Bender encontra-se aprovado para uso dos psicólogos de acordo com a Resolução 02/2003 do Conselho Federal de Psicologia: o Sistema de Pontuação Gradual – B-SPG (Sisto et al., 2005), desenvolvido para a avaliação infantil. Dessa forma, fica clara a importância e a relevância desta pesquisa que visa oferecer aos psicólogos uma versão mais atual do teste de Bender, instrumento amplamente conhecido e estudado no Brasil, e que se destina tanto a crianças como a adultos.

II. OBJETIVOS

1 – OBJETIVO GERAL

Obter os parâmetros psicométricos de normas, precisão e validade do Teste Bender Koppitz-2 para uma amostra de crianças brasileiras.

2 – OBJETIVOS ESPECÍFICOS

- a) Determinar a existência de diferenças nos resultados do Koppitz-2 em função do sexo, da idade e da escolaridade.
- b) Investigar a Validade do Bender Koppitz-2:
 1. Comparar as médias da Escala de Desenvolvimento do Bender Koppitz-2 entre as diferentes faixas etárias (validade de desenvolvimento);
 2. Verificar a relação entre os resultados da Escala de Desenvolvimento do Koppitz-2 e o Teste de Figura Complexa de Rey (Oliveira & Rigoni, 2010), para este último será considerada somente a fase de cópia.
 3. Verificar a relação entre os resultados da Escala de Desenvolvimento do Koppitz-2 e a inteligência avaliada pelo R-2: Teste Não Verbal de Inteligência Infantil (Rosa & Alves, 2012a).
 4. Verificar a relação do Koppitz-2 com o sistema de correção Koppitz (1963/1989) original.
- c) Investigar a fidedignidade do Bender Koppitz-2:
 1. Método do teste-reteste: com um intervalo de aproximadamente 30 dias entre as aplicações (média de 35,7 dias, variando entre 27 e 54 dias).

2. Precisão entre avaliadores: para verificar a precisão dos critérios adotados para a avaliação dos protocolos.
 3. Consistência interna: pelo cálculo do Alfa de Cronbach.
- d) Realizar a análise de itens do Koppitz-2 por meio do cálculo da Correlação Item-total (r_{it}) e das análises e comparações das frequências de pontuação em função da idade dos participantes.
- e) Elaborar normas para o teste considerando o sexo, a idade, a escolaridade dos participantes da pesquisa, depois de verificar a influência dessas variáveis.

III. MÉTODO

1 – Participantes

A amostra foi constituída por 623 participantes, alunos de escolas públicas de diferentes bairros da cidade de São Paulo, sendo 51,7% do sexo feminino e 48,3% do masculino. As idades variaram entre 6 e 12 anos com média de 9,05 anos e desvio padrão 1,86, que foram divididas em sete grupos com amplitude de um ano. Não foi observada diferença estatisticamente significativa entre as médias de idade dos grupos em função do sexo (meninas, média de 9,06 anos e DP=1,80; meninos, média de 9 anos e DP=1,93; $t=-0,022$; $p>0,01$). No que diz respeito a lateralidade, 90,9% eram destros e 9,1% canhotos. A escolaridade variou entre o primeiro e o nono ano do Ensino Fundamental. A Tabela 10 mostra a distribuição dos participantes em função do sexo e da idade.

Tabela 10. Distribuição de frequência da idade em função do sexo e da amostra total

Idades	Feminino		Masculino		Total	
	N	%	N	%	N	%
6	31	9,6	37	12,3	68	10,9
7	45	14,0	40	13,3	85	13,6
8	50	15,5	52	17,3	102	16,4
9	55	17,1	36	12,0	91	14,6
10	68	21,1	51	16,9	119	19,1
11	39	12,1	48	15,9	87	14,0
12	34	10,6	37	12,3	71	11,4
Total	322	51,7	301	48,3	623	100,0

Pode-se verificar pela análise da Tabela 10 que, em relação à amostra total, ainda que 10 anos tenha sido a idade com maior frequência (19,1%), os grupos etários de 7, 8, 9 e 11 anos tiveram frequências muito próximas (em torno de 15% para cada grupo aproximadamente). Os participantes de 6 anos constituíram o menor grupo (10,9%), seguido pelo de 12 anos (11,4%),

Quando se considera a amostra em função do sexo, pode-se verificar que a distribuição das idades dos meninos foi bastante equilibrada, sendo que o grupo de oito anos foi o mais frequente (17,3%) seguido pelo de 10 anos (16,9%). Já para as

meninas, o grupo maior foi o de 10 anos (21,1%), seguido pelo de 9 anos (17,1%) e foram menos frequentes as idades de 12 (10,6%) e 6 anos (9,6%).

A Tabela 11 apresenta a distribuição dos participantes em função da escolaridade e do sexo.

Tabela 11. Distribuição da escolaridade em função do sexo e da amostra total

Ano escolar	Feminino		Masculino		Total	
	N	%	N	%	N	%
1	58	18,0	62	20,6	120	19,3
2	35	10,9	34	11,3	69	11,1
3	66	20,5	52	17,3	118	18,9
4	64	19,9	49	16,3	113	18,1
5	53	16,5	49	16,3	102	16,4
6	25	7,8	33	11,0	58	9,3
7	16	5,0	17	5,6	33	5,3
8	3	0,9	3	1,0	6	1,0
9	2	0,6	2	0,7	4	0,6
Total	322	51,7	301	48,3	623	100,0

A maior parte da amostra total pertencia ao grupo com primeiro ano (19,3%). O mesmo aconteceu com o grupo masculino (20,6%), mas não com o feminino em que o grupo maior foi o do terceiro ano (20,5%). De forma geral pode-se dizer que a amostra está balanceada em relação à escolaridade, quanto ao sexo, mostrando uma distribuição percentual muito próxima. Na Tabela 12 pode-se verificar a distribuição dos participantes de acordo com a idade e o ano escolar.

Tabela 12. Distribuição de frequência absoluta e percentual da escolaridade em função da idade

Idade	6		7		8		9		10		11		12	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%
1	64	94,1	56	65,9	-	-	-	-	-	-	-	-	-	-
2	4	5,9	29	34,1	35	34,3	1	1,1	-	-	-	-	-	-
3	-	-	-	-	63	61,8	50	54,9	4	3,4	1	1,1	-	-
4	-	-	-	-	4	3,9	38	41,8	64	53,8	5	5,7	2	2,8
5	-	-	-	-	-	-	2	2,2	47	39,5	49	56,3	4	5,6
6	-	-	-	-	-	-	-	-	4	3,4	28	32,2	26	36,6
7	-	-	-	-	-	-	-	-	-	-	1	1,1	32	45,1
8	-	-	-	-	-	-	-	-	-	-	1	1,1	5	7,0
9	-	-	-	-	-	-	-	-	-	-	2	2,3	2	2,8

Pode-se verificar que para a idade de seis e sete anos há uma predominância de crianças no primeiro ano escolar (respectivamente 94,1% e 65,9%), para as de oito e nove o terceiro ano (61,8% e 54,9%), para o grupo de 10, o quarto ano (53,8%) para os de 11 o quinto (56,3%) e os de 12 o sétimo ano (45,1%). Portanto a maior parte da amostra tem escolaridade compatível com a idade. Apenas as idades de 11 e 12 anos mostram uma maior dispersão em relação à escolaridade.

Tabela 13. Distribuição de frequência da idade e sexo em função do tipo de escola

Tipo	Estadual		Municipal	
Idade	N	%	N	%
6	38	9,8	30	12,7
7	47	12,1	38	16,1
8	61	15,8	41	17,4
9	68	17,6	23	9,7
10	83	21,4	36	15,3
11	48	12,4	39	16,5
12	42	10,9	29	12,3
Sexo				
Feminino	197	50,9	125	53,0
Masculino	190	49,1	111	47,0
Total	387	62,1	236	37,9

Em relação ao tipo de escola pode-se observar uma predominância de alunos de escolas estaduais (62,1%) em relação às municipais (37,9%). A distribuição dos grupos em função dos sexos foi semelhante. O mesmo ocorreu com as idades para os dois tipos de escola, com exceção dos nove anos em que a frequência de alunos das escolas municipais foi menor quando comparada com as estaduais (9,7% e 17,6% respectivamente).

Para verificar a relação entre os pontos do Koppitz-2 (Reynolds, 2007) e as Figuras Complexas de Rey (Oliveira & Rigoni, 2010), foram avaliados 100 casos, que foram incluídos na amostra total, sendo 49% da faixa de 6 a 7 anos e 51% da faixa de 10 a 11 anos, sendo 52% meninas e 48% meninos. A escolha destes grupos etários teve como objetivo abranger as duas formas de aplicação e avaliação propostas em Koppitz-2. A Figura B do Teste de Figuras Complexas de Rey foi aplicada ao grupo mais novo e a Figura A ao mais velho. A média das idades desta amostra foi de 8,62 anos e o desvio padrão de 1,98, sendo que a diferença entre as médias de idades dos dois sexos não foi estatisticamente significativa (meninas 8,71

anos, DP=2,01 e meninos média 8,52 anos, DP=1,95; $t=0,480$; $p>0,01$). Em relação à lateralidade 86% eram destros e 14% canhotos. No que diz respeito ao tipo de escola, 52% eram alunos de escolas estaduais e 48% municipais. A Tabela 14 mostra a distribuição dos participantes em função da idade e do sexo.

Tabela 14. Distribuição de frequência por idade e sexo da amostra de validade com as Figuras Complexas de Rey

Sexo	Feminino		Masculino		Total		
	Idades	N	%	N	%	N	%
	6	10	19,2	7	14,6	17	17,0
	7	14	26,9	18	37,5	32	32,0
	10	14	26,9	12	25,0	26	26,0
	11	14	26,9	11	22,9	25	25,0
	Total	52	52,0	48	48,0	100	100,0

Na Tabela 14 observa-se que na amostra de validade com o Teste Figuras Complexas de Rey ocorreu uma predominância de crianças com 7 anos (32%) seguido pelo grupo de 10 anos (26%) e o grupo com o menor número de participantes foi o de seis anos (17%). Para o sexo feminino ocorreu uma distribuição balanceada quanto à idade, com exceção do grupo de seis anos em que a frequência foi menor. Para os meninos a idade de sete anos foi mais frequente (37,5%), seguida pelas de 10 (25%) e 11 anos (22,9%). A Tabela 15 traz a distribuição dos participantes em função da escolaridade e do sexo.

Tabela 15. Distribuição por escolaridade e sexo da amostra de validade com as Figuras Complexas de Rey

Sexo	Feminino		Masculino		Total		
	Ano escolar	N	%	N	%	N	%
	1	21	40,4	17	35,4	38	38,0
	2	3	5,8	8	16,7	11	11,0
	4	6	11,5	7	14,6	13	13,0
	5	22	42,3	16	33,3	38	38,0
	Total	52	52,0	48	48,0	100	100,0

Na Tabela 15 é encontrada uma predominância de participantes cursando o primeiro e o quinto anos tanto na amostra total (38% para cada grupo), quanto para os grupos em função do sexo (para as meninas 40,4% e 42,3% respectivamente para o primeiro e quinto ano e para os meninos 35,4% e 33,3%).

Para a validade com o Teste R-2 (Rosa & Alves, 2012a), participaram da amostra 105 crianças avaliadas pelo Koppitz-2 e pelo R-2, que foram incluídas na amostra total, sendo 49,5% de 6 e 7 anos e 50,5% de 10 e 11 anos. Sobre o tipo de escola 51,9% eram estudantes de escolas estaduais e 48,1% de municipais. A média das idades foi de 8,47 anos e o desvio padrão de 1,92. Em relação ao sexo, 50,5% eram meninas e 49,5% meninos, e a diferença entre as médias de idades dos dois grupos não foi estatisticamente significativa (média das meninas 8,42 anos, DP=1,81 e dos meninos média 8,52 anos, DP=2,04; $t=-0,276$; $p>0,05$). No que diz respeito à lateralidade, 93,3% eram destros e 6,7% canhotos. A Tabela 16 apresenta a distribuição do grupo em relação à idade e sexo.

Tabela 16. Distribuição de frequência por idade e sexo da amostra de validade do R-2

Sexo	Feminino		Masculino		Total	
	N	%	N	%	N	%
Idades						
6	8	15,1	13	25,0	21	20,0
7	19	35,8	12	23,1	31	29,5
10	21	39,6	16	30,8	37	35,3
11	5	9,4	11	21,2	16	15,2
Total	53	50,5	52	49,5	105	100,0

A Tabela 16 mostra um predomínio de crianças de 10 anos de idade desta amostra (35,3%), seguidas pelas de 7 anos (29,5%). Os meninos apresentaram uma distribuição mais equilibrada em relação às idades, já as meninas tiveram uma predominância de 10 (39,6%) e 7 anos (35,8%) tal como ocorreu com a amostra total. A Tabela 17 mostra a distribuição de frequência em função do sexo e do ano escolar da amostra de validade com o R-2.

Tabela 17. Distribuição por escolaridade e sexo da amostra de validade com o R-2

Sexo	Feminino		Masculino		Total	
	N	%	N	%	N	%
Ano escolar						
1	20	37,7	23	44,2	43	41,0
2	7	13,2	2	3,8	9	8,6
4	15	28,3	5	9,6	20	19,0
5	11	20,8	22	42,3	33	31,4

Nesta amostra pode-se constatar uma predominância de crianças no primeiro ano (41,0%), seguido pelo de quinto (31,4%), o que também ocorreu no grupo masculino, mas não no feminino em que as escolaridades mais frequentes foram o primeiro (37,7%) e o quarto anos (28,3%).

A amostra de Teste-Reteste foi constituída por 101 crianças, que também fizeram parte da amostra total, das quais 51,5% eram meninas e 48,5% meninos. Quanto à idade, 50,5% tinham de 6 a 7 anos e 49,5% de 10 a 11 anos, a média das idades foi de 8,38 anos e o desvio padrão de 1,91. Foram comparadas as médias das idades dos grupos quanto ao sexo e a diferença não foi estatisticamente significativa (a média das meninas foi 8,42 anos, DP=1,88 e dos meninos, média 8,33 anos, DP=1,96; $t=0,252$; $p>0,01$). Em relação à dominância lateral, 91,1% eram destros e 8,9 canhotos. No que diz respeito ao tipo de escola, 87,1% estudavam em escolas estaduais e 12,9% em municipais. A Tabela 18 apresenta a distribuição da amostra de teste e reteste em função do sexo e da idade.

Tabela 18. Distribuição de frequência em função da idade e sexo da amostra do Reteste

Sexo	Feminino		Masculino		Total	
	N	%	N	%	N	%
Idades						
6	10	19,2	13	26,5	23	22,8
7	16	30,8	12	24,5	28	27,7
10	20	38,5	18	36,7	38	37,6
11	6	11,5	6	12,2	12	11,9
Total	52	51,5	49	48,5	101	100,0

É possível perceber a predominância de 10 anos, tanto na amostra total (37,6%) quanto em função do sexo (38,5% para as meninas e 36,7% para os meninos). Nos dois grupos a frequência menor foi a de 11 anos. A tabela indica também que a distribuição das idades está equilibrada entre os sexos. A Tabela 19 mostra a distribuição da escolaridade da amostra do reteste em função do sexo e para a amostra total. A maior frequência foi a do primeiro ano tanto para a amostra total (42,6%) quanto em função do sexo (respectivamente 40,4% e 44,9% para as meninas e os meninos).

Tabela 19. Distribuição por escolaridade e sexo da amostra do Reteste

Sexo	Feminino		Masculino		Total	
	N	%	N	%	N	%
Ano escolar						
1	21	40,4	22	44,9	43	42,6
2	5	9,6	3	6,1	8	7,9
4	14	26,9	7	14,3	21	20,8
5	12	23,1	17	34,7	29	28,7

A precisão entre avaliadores foi realizada por meio da comparação das avaliações de dois juízes especialistas na avaliação do teste, treinados no Sistema do Bender Koppitz-2 (Reynolds, 2007). A amostra foi constituída por 62 crianças, sendo 31 de 6 e 7 anos e 31 de 8 a 12 anos, uma vez que os sistemas de pontuação e os cartões utilizados para estes dois grupos etários são diferentes. A Tabela 20 apresenta a distribuição da amostra de precisão de juízes em função do sexo e da idade.

Tabela 20. Distribuição de frequência em função da idade e sexo da amostra de juízes

Sexo	Feminino		Masculino		Total	
	N	%	N	%	N	%
Idades						
6	5	13,5	9	36,0	14	22,6
7	12	32,4	5	20,0	17	27,4
8	4	10,8	2	8,0	6	9,7
9	5	13,5	2	8,0	7	11,3
10	5	13,5	1	4,0	6	9,7
11	4	10,8	2	8,0	6	9,7
12	2	5,4	4	16,0	6	9,7
Total	37	59,7	25	40,3	62	100,0

Percebe-se que a distribuição das idades da precisão entre juízes teve um total de 31 crianças em cada uma das faixas de 6 e 7 anos e de 8 a 12 anos. A idade com menor frequência entre os meninos foi a de 10 anos com apenas um participante e entre as meninas foi de 12 anos, com apenas 2 crianças. A Tabela 21 traz a distribuição da escolaridade para a mostra do estudo de juízes, onde se pode verificar uma predominância de participantes de primeiro ano (41,9%) e na sequência o de terceiro ano (12,9%) característica também presente para os grupos em função do sexo.

Tabela 21. Distribuição da escolaridade em função do ano escolar e sexo da amostra de juízes

Sexo	Feminino		Masculino		Total	
	N	%	N	%	N	%
1	13	35,1	13	52,0	26	41,9
2	4	10,8	1	4,0	5	8,1
3	5	13,5	3	12,0	8	12,9
4	4	10,8	3	12,0	7	11,3
5	6	16,2	1	4,0	7	11,3
6	3	8,1	-	-	3	4,8
7	2	5,4	3	12,0	5	8,1
9	-	-	1	4,0	1	1,6
Total	37	59,7	25	40,3	62	100,0

A Tabela 22 mostra a distribuição da idade em função do sexo do grupo que participou do estudo que procurou verificar a relação entre o sistema Koppitz original e o Koppitz-2.

Tabela 22. Distribuição de frequência em função da idade e sexo da amostra Koppitz (1963) e Koppitz-2

Sexo	Feminino		Masculino		Total	
	N	%	N	%	N	%
6	15	9,9	19	12,8	34	11,3
7	19	12,6	20	13,4	39	13,0
8	22	14,6	21	14,1	43	14,3
9	31	20,5	21	14,1	52	17,3
10	36	23,8	27	18,1	63	21,0
11	15	9,9	21	14,1	36	12,0
12	13	8,6	20	13,4	33	11,0
Total	151	50,3	149	49,7	300	100,0

Pode-se verificar que houve uma predominância da idade de 10 anos (21,0%), seguida pela de 9 anos (17,3%), com distribuição semelhante para as demais idades para os dois sexos.

A distribuição em função da escolaridade e do sexo para a amostra da relação entre o sistema Koppitz original e o Koppitz-2 pode ser observada na Tabela 23.

Tabela 23. Distribuição da escolaridade em função do sexo da amostra Koppitz (1963) e Koppitz-2

Sexo	Feminino		Masculino		Total	
	N	%	N	%	N	%
1	22	14,6	30	20,1	52	17,3
2	19	12,6	16	10,7	35	11,7
3	38	25,2	28	18,8	66	22,0
4	34	22,5	20	13,4	54	18,0
5	20	13,2	21	14,1	41	13,7
6	14	9,3	26	17,4	40	13,3
7	4	2,6	8	5,4	12	4,0
Total	151	50,3	149	49,7	300	100,0

A maior frequência foi do terceiro ano (22,0%) seguido pelo de quarto ano (18,0%) para a amostra total, característica também verificada entre as meninas. Para os meninos o maior grupo foi o de primeiro ano (20,1%) seguido pelo terceiro ano (18,8%). A escolaridade com menor frequência foi a de sétimo ano (4,0%).

2 – Instrumentos

Para esta pesquisa foram utilizados os seguintes instrumentos: Termo de Consentimento, Teste Bender Koppitz-2, Teste das Figuras Complexas de Rey e Teste R-2 de Inteligência Não Verbal para Crianças.

2.1 – Termo de consentimento

O termo de consentimento assinado pelos pais é apresentado no Anexo I.

2.2 – Teste Bender Koppitz-2

2.3 – Teste das Figuras Complexas de Rey

As Figuras Complexas de Rey foram elaboradas por André Rey em 1942 com a finalidade inicial de auxiliar no “*diagnóstico diferencial entre a debilidade mental constitucional e o déficit adquirido em consequência de traumatismo crânio-cerebral*” (Oliveira, 1999, p.9). As figuras foram criadas considerando os seguintes critérios: não apresentar uma significação clara; ser de fácil construção gráfica e ter uma estrutura suficientemente complicada e que exige uma atividade perceptiva analítica e organizadora para a sua reprodução.

O teste é composto por duas Figuras: A Figura “A” mais complexa, que deve ser utilizada a partir dos 5 anos e a Figura “B” mais simples, que deve ser administrada somente a crianças com idades entre 4 e 7 anos. Neste trabalho serão utilizadas as Figuras A e B em função das características da amostra estudada (entre 6 e 12 anos). Embora a Figura A possa ser aplicada a partir dos 5 anos, por se tratar de uma figura mais complexa do que a B, optou-se por utilizar a última com os participantes até 7 anos, por ser mais adequada para essa faixa etária. A aplicação é individual e pode ser usado na faixa entre 4 e 88 anos de idade. Não há limite de tempo de aplicação, mas esta costuma variar entre 5 e 25 minutos.

Oliveira e Rigoni (2010) afirmam que, ao aplicar as Figuras Complexas de Rey tanto em sua fase de cópia quanto na de memória, o objetivo é “*avaliar as funções neuropsicológicas de percepção visual e memória imediata*” (p. 15), ou seja, verificar como a pessoa apreende as informações e os dados perceptivos, que lhe são apresentados, e o que consegue conservar espontaneamente na memória. As autoras apontam ainda que as técnicas psicológicas, que envolvem cópias de figuras, são em geral bem aceitas pelas pessoas, além de apresentarem um baixo custo por necessitarem apenas de lápis e papel.

As instruções de aplicação e avaliação são diferentes para cada Figura. Para a Figura A, a tarefa é primeiramente solicitar a cópia da figura apresentada, informando que durante sua execução serão trocadas as cores dos lápis de forma que ao final a cópia ficará colorida. O aplicador deverá evitar que sejam executadas muitas partes da figura com a mesma cor, devendo trocar o lápis durante a realização, lembrando que devem ser utilizadas no máximo seis cores. A finalidade da troca de cor é verificar o tipo de cópia produzida. Permite também ver a sequência das partes desenhadas e como foi o planejamento na cópia. A marcação do tempo deve ser iniciada assim que for apresentada a figura. Após a conclusão da cópia é solicitado que a figura seja reproduzida novamente de memória, usando o mesmo procedimento de troca dos lápis de cores. Em nenhuma das partes é permitido o uso de borracha.

A correção é feita em duas etapas: primeiramente deve-se identificar o tipo da cópia. Osterrieth (1945, citado por Oliveira & Rigoni, 2010) propôs uma classificação, com base na análise de diversas reproduções, que leva em conta a forma como o desenho é estruturado. Chegou a seis tipos de cópia que são organizados da “mais à menos racional”, ou seja, os tipos iniciais são os que mais se aproximam da figura

original copiada e os finais aqueles que mais se distanciam dela. Neste sistema de classificação são considerados o desenvolvimento e a percepção visual que direciona o trabalho de cópia (Rey, 1959, citado por Oliveira & Rigoni, 2010).

Na segunda etapa da correção são atribuídos pontos para a reprodução adequada das partes da figura. A Figura A foi dividida em 18 unidades que são pontuadas de acordo com a precisão com a qual foram realizadas. Dessa forma cada unidade receberá uma pontuação entre 0 e 2 pontos, envolvendo a qualidade e a localização adequada da forma: reproduzida corretamente e bem situada, 2 pontos; mal situada, 1 ponto; deformada ou incompleta, mas reconhecível, bem situada 1 ponto; mal situada 0,5 pontos; irreconhecível ou ausente: 0 ponto. A pontuação máxima em cada fase é de 36 pontos. Esses critérios de avaliação são usados, tanto na fase de cópia, quanto na de memória.

O manual do teste (Oliveira & Rigoni, 2010) apresenta os estudos brasileiros de Precisão (teste-reteste, consistência interna pelo Alfa de Cronbach e correlação entre avaliadores), Validade (dois tipos: convergente, com os Subtestes de Dígitos e Aritmética do WISC-III, e de critério pela comparação de resultados com grupos clínicos: adolescentes usuários de drogas, adultos alcoolistas e epiléticos) e Normas, que são expressas em percentis para a pontuação e frequência para o tipo de cópia realizada. As tabelas são separadas em função da idade, considerada ano a ano entre 5 a 15 anos, e posteriormente por faixas de idade (16 a 20; 21 a 40; 41 a 60 e 61 anos ou mais). São apresentadas tabelas para o Tipo de Cópia (com a frequência de reprodução de cada um dos tipos descritos acima), para o Total de Pontos e o Tempo utilizado para a reprodução das figuras (em percentis, sendo que para estas duas últimas, tanto na fase de cópia quanto na de memória).

Para a Figura B deve-se apresentar a figura à criança e pedir que ela a copie a lápis da melhor forma que conseguir. Marca-se o tempo assim que ela iniciar a cópia. Após no máximo três minutos da conclusão da cópia, solicita-se que a criança reproduza a figura novamente de memória. Os lápis de cor também devem ser trocados.

O manual do Rey traz os estudos de Precisão (consistência interna pelo Alfa de Cronbach), de Validade (três estudos: convergente com o teste de Avaliação Neurológica Infantil (ANI), que avalia o desempenho psicomotor, e pela correlação entre os pontos das Figuras A e B aplicadas na mesma criança; de critério com um

grupo clínico de crianças com epilepsia). As normas foram elaboradas em tabelas de percentis em função da idade para cada ano na faixa entre 4 e 7 anos.

2.4 – R-2: Teste Não Verbal de Inteligência para Crianças

O R-2 - Teste Não Verbal de Inteligência para Crianças (Rosa & Alves, 2000, 2012a) foi elaborado por Rynaldo de Oliveira e tem como objetivo avaliar o fator *g* de Spearman, tendo como base o *Teste de Matrizes Progressivas Coloridas de Raven* (Raven, 1947).

Spearman foi o primeiro que se preocupou em investigar, se as razões da variabilidade do rendimento cognitivo eram motivadas pela existência de uma única capacidade intelectual ou fruto de múltiplas capacidades (Andrés-Pueyo, 2006). Segundo esta teoria, a inteligência é composta por dois fatores: um fator geral (fator *g* de *general*) e diversos fatores específicos *s* (de *specific*). Sendo assim, ao se resolver qualquer tipo de situação ou em qualquer tipo de atividade intelectual, estes dois fatores estariam presentes (Bernstein, 1955b; Schelini, 2002; Silva, 2003; Spearman, 1955).

Para Spearman (1955), o fator geral seria comum a todas as capacidades humanas e os específicos estariam diretamente relacionados à capacidade específica em questão, ou seja, à atividade mental que está sendo utilizada para a tarefa. Dessa forma, pode-se afirmar que nem sempre estes fatores irão desempenhar o mesmo papel: em algumas situações o fator *g* pode ser o principal, em outras, entretanto, os fatores específicos podem ocupar este lugar. O fator *g* recebeu este nome “*porque, ainda que varie livremente de um indivíduo para os outros, permanece inalterável em um mesmo indivíduo, em relação a todas as demais capacidades correlacionadas*” (Spearman, 1955, p.82). O fator *s* por sua vez pode variar entre indivíduos diferentes e também para diferentes capacidades de um mesmo indivíduo.

O Teste R-2 é derivado do Teste R-1 (Alves, 2002; Oliveira, 1973) que é um teste de inteligência não verbal para adultos, empregado em avaliações para diversos fins, principalmente para a obtenção da Carteira Nacional de Habilitação (CNH). Segundo Oliveira (1997), o R-2 “*apresenta uma variedade de tarefas que se pressupõe serem uma amostragem adequada das funções intelectuais mais importantes*” (p. 5).

O R-2 é constituído por um conjunto de 30 pranchas com figuras coloridas (objetos concretos ou figuras abstratas) e que estão organizadas em ordem crescente de dificuldade. A construção do instrumento segue os mesmos princípios do Teste das Matrizes Progressivas de Raven, ou seja, os itens são formados por figuras que apresentam uma parte faltando. A tarefa dos examinandos se constitui em localizar entre as alternativas disponíveis abaixo, qual aquela que completa corretamente a figura de cima de acordo com o raciocínio envolvido em cada item. São apresentadas seis ou oito alternativas de respostas, sendo que apenas uma delas é a correta. Os primeiros são itens de complementação de figura concreta, mas durante o teste são exigidos outros tipos de raciocínio para a resolução dos itens (identidade, analogia, raciocínio classificatório, completamento de fundo, alternância de figuras, raciocínio numérico e espacial, entre outros).

A aplicação é individual e o examinador anota a resposta da criança em cada um dos itens. O teste é destinado a crianças entre 4 anos e 9 meses e 11 anos e 8 meses. Não há limite de tempo para a aplicação uma vez que é um teste de capacidade/poder, no entanto, o tempo médio de aplicação é de 8 minutos e a maior parte das crianças gastam entre 5 e 15 minutos.

A correção é feita com um crivo com as respostas corretas. O total de pontos é obtido em função do número de acertos que devem ser convertidos em percentis. Rosa e Alves (2012a) propõem também uma análise qualitativa relativa aos tipos de erros cometidos e que permite verificar entre os raciocínios exigidos para resolver cada um dos itens, aquele que a criança teve maior dificuldade. O manual do teste (Rosa & Alves, 2000, 2012a) apresenta os estudos psicométricos de precisão, validade que sustentam a qualidade do instrumento para a avaliação da inteligência, além de normas para São Paulo, Assis e Recife.

3 – Procedimento

Antes de ser iniciada a pesquisa o projeto foi submetido à avaliação do Comitê de Ética em Pesquisa (CEP) do Instituto de Psicologia da Universidade de São Paulo, que aprovou¹⁰ sua realização. Os participantes foram provenientes de 12 escolas estaduais e 9 municipais de diferentes regiões da cidade de São Paulo, a

¹⁰ Certificado de Apresentação para Apreciação Ética – CAAE nº 22355814.7.0000.5561, parecer nº 569.199.

fim de possibilitar certa heterogeneidade da amostra. Antes da coleta de dados foram feitos contatos com a direção das escolas a fim de apresentar os objetivos da pesquisa, bem como explicar as atividades que seriam realizadas pelas crianças e também foi informado a elas que poderiam decidir se queriam ou não participar da pesquisa.

As crianças foram sorteadas em função do número de alunos das salas de aula e de forma a equilibrar a amostra em função do sexo, idade e escolaridade dos participantes. Foram entregues os termos de consentimento e a solicitação de autorização dos pais (ainda que em função do tipo de pesquisa proposto, o Comitê de Ética tenha autorizado a realização da pesquisa, mesmo sem a assinatura destes documentos) e somente foram submetidas aos testes as crianças que devolveram os termos assinados. A aplicação foi feita durante o período de aula: as crianças eram retiradas da sala para a realização das aplicações nos locais disponibilizados pelas escolas. As salas ofereciam condições adequadas de acomodação, iluminação e ausência de estímulos ou interferências externas que pudessem prejudicar as aplicações e conseqüentemente os resultados. Os testes foram aplicados por 8 psicólogos formados e 5 estudantes de Psicologia que foram devidamente treinados na aplicação, sendo que os estudantes estavam o tempo todo sob supervisão de um psicólogo.

O Bender Koppitz-2 foi aplicado individualmente de acordo com as instruções descritas no manual (Reynolds, 2007) que foram traduzidas e são apresentadas no Anexo II. Foram aplicados os cartões 1 a 13 nos participantes até 7 anos e 5 a 16, nos de 8 anos ou mais. Foi elaborada uma folha de registro e solicitado aos aplicadores que preenchessem os dados das crianças e da aplicação, incluindo a ordem de realização das partes de cada um dos desenhos e os comportamentos, comentários e atitudes das crianças durante o teste. A folha de registro usada pelo examinador para anotar os comportamentos das crianças entre 5 e 7 anos está no Anexo III e para as participantes de 8 anos ou mais no Anexo IV. Antes das aplicações foi estabelecido um *rapport* a fim de proporcionar uma relação adequada para a realização dos testes.

Os protocolos foram avaliados pelo pesquisador e por cinco psicólogos treinados por ele, seguindo os critérios de avaliação do Koppitz-2 propostos por Reynolds (2007) em função da idade do participante. Os desenhos foram pontuados em função da qualidade de sua reprodução sendo atribuídos escores um ou zero

para cada um dos itens. A pontuação máxima para as crianças menores é de 34 pontos e, para as maiores, de 45 pontos.

Para investigar a validade entre o R-2 e o Koppitz-2, os testes foram aplicados no mesmo dia um após o outro, sendo que 48,6% dos participantes fizeram primeiro o R-2 e o restante do grupo (51,4%), primeiro o Koppitz-2. O Teste R-2 foi corrigido obtendo-se o total de acertos (no máximo 30 pontos).

O Teste das Figuras Complexas de Rey e o Bender Koppitz-2 também foram aplicados no mesmo dia, um após o outro, sendo que 52% das crianças fizeram primeiro o teste de Rey e o restante do grupo (48%) primeiro o Bender. As Figuras Complexas de Rey foram avaliadas conforme as orientações do manual, sendo a pontuação máxima para a Figura A, 36 pontos e para a Figura B, 31 pontos. Somente a fase de cópia das Figuras Complexas de Rey foi aplicada neste estudo.

Para investigar a relação do sistema Koppitz-2 com os Indicadores Maturacionais de Koppitz (1963/1989) foram selecionados 300 protocolos da amostra total, que também foram avaliados pelos dois sistemas, por um psicólogo especialista nesta avaliação. No sistema de Koppitz são pontuadas as características distorção da forma, integração, rotação e perseveração. A pontuação máxima é de 30 pontos e quanto maior o total de pontos, pior é o resultado no teste.

Para a precisão teste-reteste, foram avaliadas 101 crianças da amostra total, sendo 50,5% entre 6 e 7 anos e 49,5% entre 10 e 11. Elas foram submetidas ao teste duas vezes, com um intervalo médio entre as aplicações de aproximadamente 35,7 dias (variando entre 27 e 54 dias).

Em relação à precisão entre avaliadores, foram extraídos da amostra total 62 protocolos, metade de crianças mais novas (6 e 7 anos) e metade de mais velhas (8 a 12 anos), que foram avaliados por dois juízes especialistas em Bender e treinados nos sistemas de avaliação do Koppitz-2 (Reynolds, 2007).

Todos os testes foram avaliados segundo as instruções dos respectivos manuais. Os dados foram tabulados e os tratamentos estatísticos efetuados com o Programa Estatístico SPSS 21.0.

4 – Tratamento Estatístico

Para verificar a relação entre a idade e os resultados no Koppitz-2 foram calculadas as Correlações de Pearson entre os totais de pontos no teste e as idades

dos participantes. Também foram realizadas comparações entre as médias dos grupos em função do sexo, idade, escolaridade e tipo de escola por meio da Análise de Variância para determinar os efeitos de cada uma dessas variáveis no total de pontos.

A relação entre o Koppitz-2 e os Testes R-2 e Figuras Complexas de Rey foi calculada pela Correlação Linear de Pearson, para verificar a validade entre testes que avaliam construtos semelhantes (no caso das Figuras Complexas de Rey – validade concorrente/convergente) e construtos relacionados (no caso do R-2).

A relação entre os dois sistemas de avaliação Koppitz-2 e o original foi investigada pela Correlação Pearson, lembrando que um sistema pontua os acertos e o outro, os erros, a correlação entre eles é negativa. As correlações foram obtidas tanto para o total de pontos de todas as figuras do Koppitz-2 para a respectiva idade, como apenas para as nove comuns às duas formas de administração do teste.

A fidedignidade Teste-Reteste investigada pela Correlação Linear de Pearson e o efeito de aprendizagem, pelo *teste t de Student* para amostras pareadas.

Para a precisão entre juízes foi obtida a Correlação Linear de Pearson dos totais de pontos das avaliações dos dois especialistas e a comparação das médias de pontos foi feita pela Análise de Variância (*One Way Anova*). Ainda para a precisão entre juízes também foi calculado o Coeficiente Kappa de Concordância, que é uma medida de concordância entre juízes e/ou avaliadores. Este coeficiente pode variar entre “0” (zero), que indica ausência de concordância e “1” (um), que significa uma completa concordância entre os juízes (Siegel & Castellan Jr, 2006). Esse coeficiente mostra o grau da concordância entre dois ou mais juízes ou entre dois métodos de classificação (por ex. dois testes), baseando-se para isso no número de resultados concordantes (Baltar & Okano, s.d.). Estes autores propõem a seguinte escala de classificação para o coeficiente Kappa: resultados entre 0 e 0,39 sugerem um fraca concordância, entre 0,40 e 0,59 moderada, entre 0,60 e 0,79 substancial e valores entre 0,80 e 1, uma concordância quase perfeita.

A fidedignidade referente à consistência interna foi verificada pelo cálculo do Alfa de Cronbach, que é um coeficiente de precisão que analisa “a congruência que cada item do teste tem com o restante dos itens do mesmo teste” (Pasquali, 2003, p. 203).

A análise de itens foi realizada com os dados da amostra total. Para verificar a discriminação dos itens foram calculadas as Correlações Item-Total (r_{it})

considerando o grupo de crianças de 6 e 7 anos e de 8 a 12 anos separadamente (uma vez que os itens avaliados não são os mesmos). Também foram analisadas as frequências de pontuações por idades e comparadas pelo Qui-Quadrado (χ^2) para identificar, se elas diferenciaram ou não os participantes em função de sua idade.

As normas para o teste foram elaboradas em Percentis, considerando as diferenças entre as médias obtidas pelas Análises de Variância e *Testes t de Student* em relação às variáveis: sexo, idade e escolaridade dos participantes.

IV. RESULTADOS

1) Diferenças entre idade, sexo e tipo de escola

Foram conduzidas análises estatísticas para verificar a existência de diferenças estatisticamente significantes entre as médias do total de pontos no Koppitz-2 em função da idade, sexo e tipo de escola dos participantes da pesquisa. Como já foi mencionado anteriormente, todas as análises serão feitas separadamente para as crianças de 6 e 7 anos e de 8 a 12 anos, uma vez que tanto a aplicação (tipo e número de figuras aplicadas) como a avaliação (o total de pontos máximo para os mais novos é de 34 pontos e para os mais velhos 45 pontos) são diferentes para estes grupos.

A Figura 1 mostra o histograma da distribuição das frequências do total de pontos do Koppitz-2 para a amostra de 6 e 7 anos (mais novos).

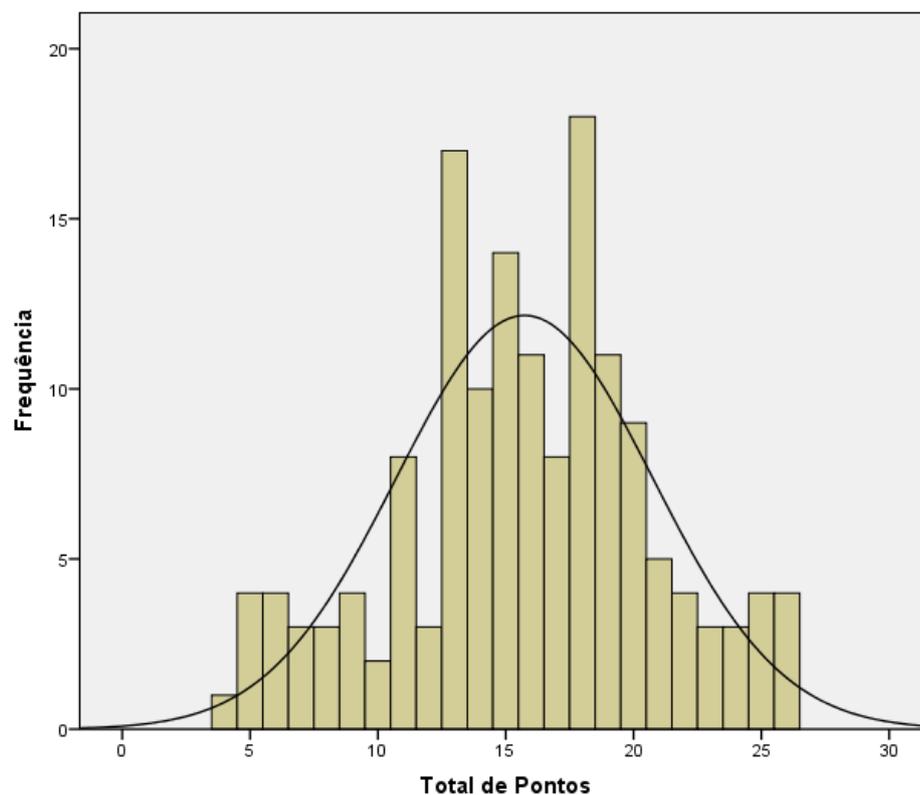


Figura 1. Histograma da distribuição do total de pontos do Koppitz-2 para o grupo dos mais novos

Pode-se observar que o histograma do total de pontos do Koppitz-2 para as crianças de 6 e 7 anos apresenta um formato muito próximo ao da curva normal. A média foi de 15,73 pontos (DP=5,02) a moda foi de 18 pontos e a mediana 16, muito próxima da média. As pontuações variaram entre 4 e 26 pontos.

Para verificar a existência de diferenças estatisticamente significantes entre as médias do total dos pontos no Koppitz-2, foi realizada a Análise de Variância Univariada (Modelo Linear Geral) considerando o sexo, idade e tipo de escola. Os resultados são apresentados na Tabela 24 para 6 e 7 anos.

Tabela 24. Análise de Variância do total de pontos do Koppitz-2 por idade, sexo e tipo de escola para 6 e 7 anos

	<i>gl</i>	<i>F</i>	<i>Sig.</i>
Modelo Corrigido	7	4,723	0,000**
Interseção	1	1632,929	0,000**
Idade	1	19,947**	0,000
Sexo	1	4,791*	0,030
Tipo de Escola	1	2,759	0,099
Interação Idade x Sexo	1	3,151	0,078
Interação Idade x Tipo de Escola	1	0,048	0,827
Interação Sexo x Tipo de Escola	1	0,219	0,640
Interação Idade x Sexo x Tipo de Escola	1	0,791	0,375
Erro	145		
Total	153		
Total Corrigido	152		

** Significante a 0,01

gl = graus de liberdade

* Significante a 0,05

Na Tabela 24, para as crianças mais novas, foram encontradas diferenças significantes entre as médias de pontos do Koppitz-2 em função da idade ($F=19,947$; $p<0,001$) e do sexo ($F=4,791$; $p<0,05$), mas não do tipo de escola ($F=2,759$; $p>0,05$) e nem para as interações entre estas variáveis.

A Tabela 25 mostra as médias e desvios padrão do total de pontos das crianças de 6 e 7 anos por idade, sexo e tipo de escola. Os resultados mostraram que a média das meninas foi maior do que a dos meninos e a das crianças de 7 anos, maior do que a de 6, o que indica uma tendência de aumento da média com a progressão da idade. A correlação entre a idade e o total de pontos foi de 0,352 (significante a 0,01). Apesar de ser uma correlação baixa, é estatisticamente significativa e positiva, o que indica uma relação entre as duas variáveis,

confirmando o que foi observado na comparação entre as médias, de que à medida que a criança fica mais velha ocorre um aumento no total de pontos. Estes resultados indicam que as médias diferenciaram os grupos em função da idade e do sexo, o que justifica a necessidade de serem elaboradas tabelas normativas separadas de acordo com estas características. Ainda que as escolas estaduais tenham apresentado média maior do que as municipais, a diferença não foi significativa, por isso não são necessárias normas para o tipo de escola.

Tabela 25. Médias e desvios padrão do total de pontos do Koppitz-2 dos mais novos

Variáveis	Subgrupos	Total	
		Média	DP
Idade (anos)	6	13,76	5,25
	7	17,31	4,24
Sexo	Feminino	16,57	4,88
	Masculino	14,91	5,05
Tipo de escola	Estadual	16,32	4,79
	Municipal	15,00	5,24
Total		15,73	5,02

A Figura 2 mostra as médias das pontuações por idade para de 6 e 7 anos.

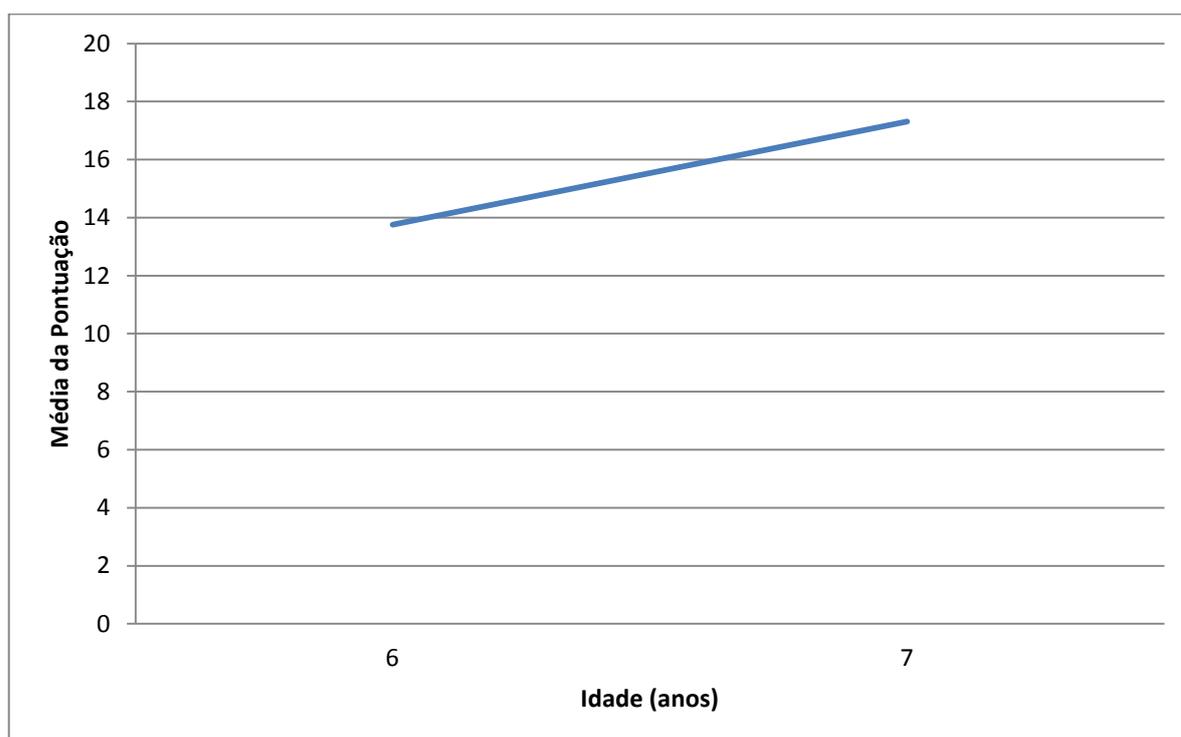


Figura 2. Médias das pontuações por idade para as crianças de 6 e 7 anos.

Também foi calculada uma Análise de Variância do tempo de aplicação por idade e sexo para os mais novos, que é mostrada na Tabela 26.

Tabela 26. Análise de Variância do tempo de aplicação do Koppitz-2 por idade e sexo para os mais novos

	<i>gl</i>	<i>F</i>	<i>Sig.</i>
Modelo Corrigido	3	2,548	0,058
Interseção	1	607,677	0,000
Idade	1	6,998**	0,009
Sexo	1	0,176	0,176
Interação Idade x Sexo	1	0,301	0,584
Erro	141		
Total	145		
Total Corrigido	144		

Significante a 0,01
gl = graus de liberdade

A Tabela 26 mostra que as médias de tempo foram diferentes ($F=6,998$; $p<0,01$) somente em função das idades. As crianças mais novas tenderam a utilizar um tempo maior para realizar os desenhos. A Tabela 27 apresenta as médias e desvios padrão dos tempos por sexo e idade.

Tabela 27. Médias e desvios padrão do tempo de aplicação do Koppitz-2 dos mais novos por idade e sexo

Variáveis	Subgrupos	Tempo	
		Média	DP
Idade (anos)	6	10,57	5,82
	7	8,50	3,27
Sexo	Feminino	9,22	4,00
	Masculino	9,63	5,29
Total		9,43	4,69

A Figura 3 mostra a distribuição das frequências do total de pontos do Koppitz-2 para o grupo de 8 a 12 anos.

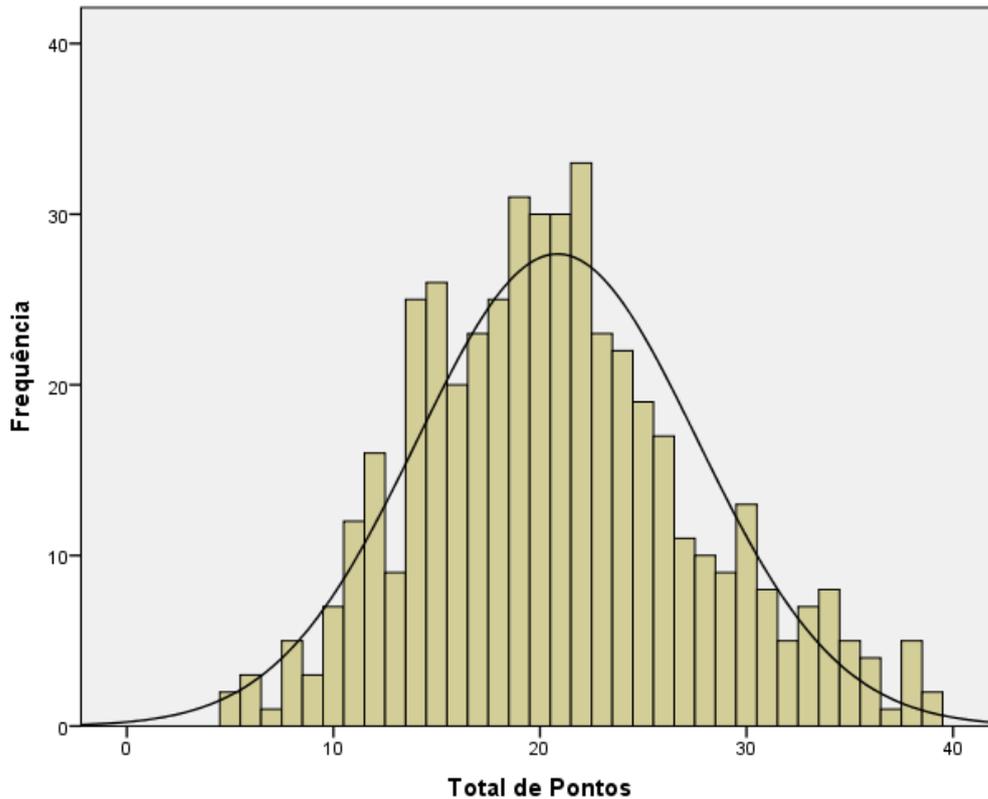


Figura 3. Histograma da distribuição do total de pontos para o grupo de 8 a 12 anos

O histograma da distribuição do total de pontos para 8 a 12 anos tem um formato muito próximo ao da curva normal. A média de pontos foi de 20,83 pontos (DP=6,80) a moda foi de 22 pontos e a mediana 20, muito próxima portanto do valor da média. Os resultados variaram entre 5 e 39 pontos.

Para verificar a existência de diferenças estatisticamente significantes entre as médias do total dos pontos para 8 a 12 anos, os resultados foram comparados por meio da Análise de Variância Univariada (Modelo Linear Geral), considerando a idade, o sexo e o tipo de escola. A Tabela 28 apresenta os resultados.

Pode-se perceber que as diferenças entre as médias foram estatisticamente significantes em função da idade ($F=5,964$; $p<0,001$) e do tipo de escola ($F=8,278$; $p<0,01$), mas não do sexo ($F=0,732$; $p>0,05$). Quanto às interações entre as variáveis, a interação entre idade e sexo teve diferenças significantes ($F=2,433$; $p<0,05$).

Tabela 28. Análise de Variância para o total de pontos do Koppitz-2 por idade, sexo e tipo de escola para 8 a 12 anos

	<i>gl</i>	<i>F</i>	<i>Sig.</i>
Modelo Corrigido	19	2,598	0,000
Interseção	1	4132,007	0,000
Idade	4	5,964**	0,000
Sexo	1	0,732	0,393
Tipo de Escola	1	8,278**	0,004
Interação Idade x Sexo	4	2,433*	0,047
Interação Idade x Tipo de Escola	4	0,346	0,847
Interação Sexo x Tipo de Escola	1	0,123	0,726
Interação Idade x Sexo x Tipo de Escola	4	1,463	0,213
Erro	450		
Total	470		
Total Corrigido	469		

** Significante a 0,01

* Significante a 0,05

gl = graus de liberdade

A Tabela 29 apresenta as médias e desvios padrão dos pontos do Koppitz-2 para 8 a 12 anos.

Tabela 29. Médias e desvios padrão do total de pontos do Koppitz-2 para 8 a 12 anos

Variáveis	Subgrupos	Total	
		Média	DP
Idade (anos)	8	18,31	6,22
	9	20,86	6,90
	10	20,91	6,28
	11	21,86	6,35
	12	23,04	7,82
Tipo de escola	Estadual	20,22	6,49
	Municipal	21,93	7,21
Sexo	Feminino	21,27	6,81
	Masculino	20,35	6,77
Total		20,83	6,80

Assim como foi observado para 6 e 7 anos, as crianças de 8 a 12 anos também mostraram uma tendência de aumento da média com o crescimento da idade. Para verificar esta relação, foi calculada a correlação entre a idade e o total de pontos do Koppitz-2. Foi obtida uma correlação baixa, ainda que estatisticamente significativa ($r=0,212$; $p=0,01$), o que aponta uma relação entre as duas variáveis, confirmando a característica observada na comparação entre as médias de que à medida que a idade cresce há um aumento no total de pontos. Uma vez que foram

observadas diferenças estatisticamente significantes nas interações entre idade e sexo, optou-se por elaborar uma tabela que apresentasse os resultados das médias e desvios padrão para os totais por sexo e idade e para os dois tipos de escola (Tabela 30).

Tabela 30. Médias e desvios padrão do total de pontos do Koppitz-2 dos mais velhos em função da idade por sexo e tipo de escola

Idade	Sexo	Tipo de Escola	Média	DP	N
8	Feminino	Total	19,50	6,24	50
	Masculino	Total	17,17	6,04	52
	Total	Estadual	17,66	5,76	61
		Municipal	19,29	6,80	41
		Total	18,31	6,22	102
9	Feminino	Total	20,04	6,72	55
	Masculino	Total	22,11	7,07	36
	Total	Estadual	20,18	6,45	68
		Municipal	22,87	7,89	23
		Total	20,86	6,90	91
10	Feminino	Total	21,84	6,48	68
	Masculino	Total	19,67	5,83	51
	Total	Estadual	20,37	5,89	83
		Municipal	22,14	7,03	36
		Total	20,91	6,28	119
11	Feminino	Total	22,82	6,71	39
	Masculino	Total	21,08	6,01	48
	Total	Estadual	21,21	6,38	48
		Municipal	22,67	6,31	39
		Total	21,86	6,35	87
12	Feminino	Total	22,97	7,86	34
	Masculino	Total	23,11	7,89	37
	Total	Estadual	22,60	7,75	42
		Municipal	23,69	8,01	29
		Total	23,04	7,82	71
Total	Feminino	Total	21,27	6,81	246
	Masculino	Total	20,35	6,77	224
	Total	Estadual	20,22	6,49	302
		Municipal	21,93	7,21	168
		Total	20,83	6,80	470

Para verificar quais grupos etários se diferenciaram foi utilizado o teste de *Post Hoc de Tukey*, os dados são apresentados na Tabela 31.

Tabela 31. Resumo do Teste de Tukey por idade para os pontos do Koppitz-2 para 8 a 12 anos

Idade	N	Subconjunto	
		1	2
8	102	18,31	
9	91	20,86	20,86
10	119		20,91
11	87		21,86
12	71		23,04
Sig.		0,062	0,167

Pode-se verificar que somente a média de 8 anos não mostrou diferença em relação a de 9 anos, mas se diferenciou de 10, 11 e 12 anos, enquanto que 9 anos não mostrou diferenças significantes em relação a nenhuma das idades. A Figura 4 mostra as médias de pontos por idade para as crianças de 8 a 12 anos.

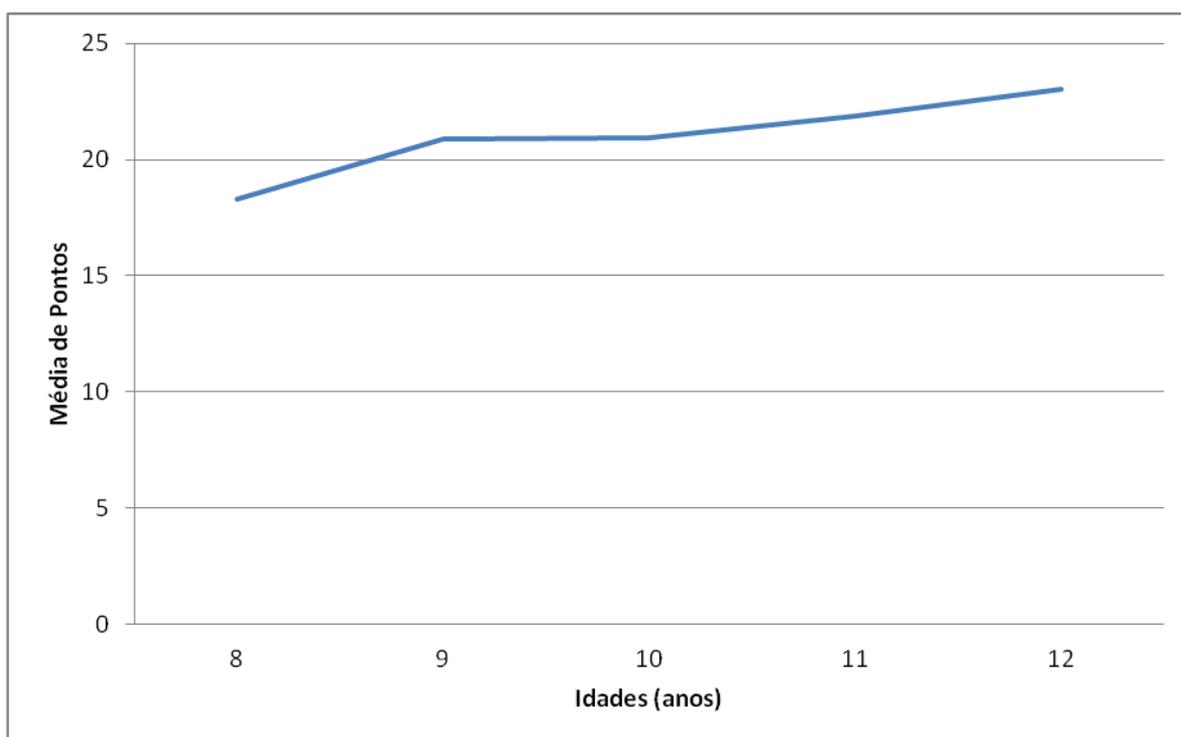


Figura 4. Médias de pontos por idade para 8 a 12 anos.

A Figura 4 demonstra o aumento das médias de pontos com a idade, sendo que a maior inclinação se deu entre 8 e 9 anos, praticamente não houve mudança entre 9 e 10 e há um aumento leve entre 11 e 12.

Uma vez que foram encontradas diferenças estatisticamente significantes entre as médias da faixa de 8 a 12 anos em função da interação idade e sexo, foram

calculados os Testes *t* para cada sexo entre as idades, para identificar quais grupos se diferenciaram. As Tabelas 32 e 33 mostram esses resultados.

Tabela 32. Médias, desvios padrão e teste t por idade para o grupo feminino de crianças de 8 a 12 anos

Idades	Média	DP	Teste t			
			9	10	11	12
8	19,50	6,24	-0,424	-1,979*	-2,410**	-2,251*
9	20,04	6,72	-	-1,508	-1,981	-1,876
10	21,84	6,48	-	-	-0,745	-0,774
11	22,82	6,71	-	-	-	-0,088
12	22,97	7,86	-	-	-	-

** Significantes a 0,01

* Significantes a 0,05

Tabela 33. Médias, desvios padrão e teste t por idade para o grupo masculino de crianças de 8 a 12 anos

Idades	Média	DP	Teste t			
			9	10	11	12
8	17,17	6,04	-3,515**	-2,132*	-3,244**	-3,884**
9	22,11	7,07	-	1,762	0,719	-0,569
10	19,67	5,83	-	-	-1,190	-2,245*
11	21,08	6,00	-	-	-	-1,344
12	23,11	7,89	-	-	-	-

** Significantes a 0,01

* Significantes a 0,05

Pode-se verificar que, para o sexo feminino, somente a média de 8 anos mostrou diferença em relação a todas as idades (exceto para 9 anos).

Em relação ao sexo masculino, 8 anos mostrou diferença de média em relação a todas as idades, e 10 anos também se diferenciou do grupo de 12anos. As demais idades não mostraram diferenças significantes entre elas.

A Tabela 28 mostra ainda que foram observadas diferenças estatisticamente significantes quando se considera o tipo de escola, sendo que a média dos alunos das escolas municipais foram maiores do que das estaduais. Esta característica foi observada somente quando se considera o efeito isolado da variável e não quando ela foi relacionada ao sexo e mesmo a idade dos participantes.

Os resultados das análises de variância indicam, portanto a necessidade de serem elaboradas tabelas de normas separadas para 8 anos e para as escolas estaduais e municipais para a faixa de 8 a 12 anos.

Assim como para a faixa de 6 e 7 anos, também foi realizada Análise de Variância do tempo de aplicação para 8 e 12 anos, em função da idade, sexo e tipo de escola. A Tabela 34 mostra estes resultados.

Tabela 34. Análise de Variância para o tempo de aplicação por idade, sexo e tipo de escola para 8 a 12 anos

	<i>gl</i>	<i>F</i>	<i>Sig.</i>
Modelo Corrigido	19	0,713	0,806
Interseção	1	1596,772	0,000
Idade	4	1,439	0,220
Sexo	1	0,049	0,825
Tipo de Escola	1	5,978*	0,015
Interação Idade e Sexo	4	0,204	0,936
Interação Idade e Tipo de Escola	4	0,227	0,923
Interação Sexo e Tipo de Escola	1	0,139	0,709
Interação Idade, Sexo e Tipo de Escola	4	0,222	0,926
Erro	419		
Total	439		
Total Corrigido	438		

Significante a 0,05

gl = graus de liberdade

A Análise de Variância indicou diferença apenas para o tipo de escola no tempo de execução dos desenhos. A Tabela 35 mostra as médias e desvios padrão por sexo, tipo de escola e idade para a faixa de 8 a 12 anos.

Tabela 35. Médias e desvios padrão do tempo de aplicação do Koppitz-2 para crianças de 8 a 12 anos

Variáveis	Subgrupos	Tempo	
		Média	DP
Sexo	Feminino	14,52	7,04
	Masculino	14,09	6,83
Tipo de escola	Estadual	13,77	6,08
	Municipal	15,30	8,19
Idade	8	14,49	7,08
	9	15,20	6,92
	10	14,80	8,06
	11	13,74	6,57
	12	13,06	5,24
Total		14,32	6,94

Somente para o tipo de escola a diferença entre as médias do tempo de aplicação foi significativa, sendo que a média dos alunos das escolas municipais foi maior.

Reynolds (2007) não apresentou no manual do Koppitz-2 as médias e os desvios padrão das pontuações da amostra de normatização, o que impossibilitou que os dados fossem comparados com os resultados da normatização brasileira para verificar se a diferença entre as médias era estatisticamente significativa. Dessa forma, optou-se por elaborar o gráfico referente ao percentil 50 para as duas amostras, pois este resultado é o que mais se aproxima da média.

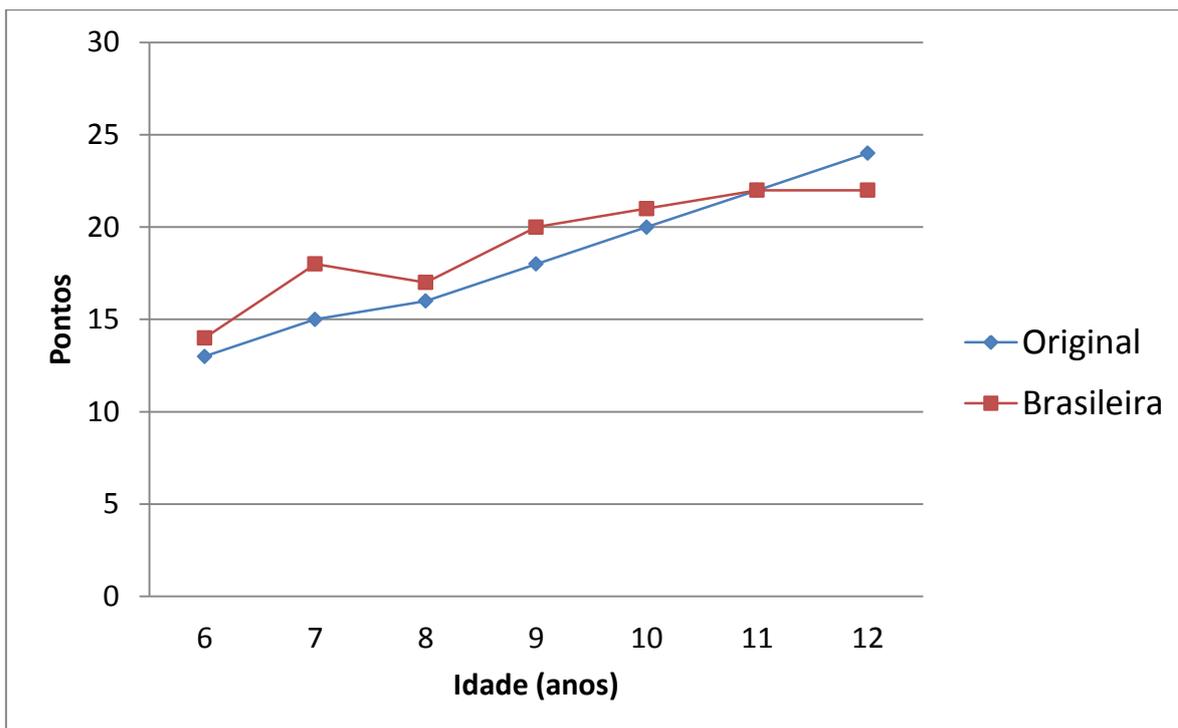


Figura 5. Comparação do percentil 50 das normas americana e brasileira

Na Figura 5 pode-se perceber que a média de pontos da amostra brasileira foi mais alta do que a americana para as idades de 6 a 10 anos, para 11 anos foi igual e para 12 anos a média americana foi maior.

2) Validade

Em linhas gerais “a validade de um teste refere-se àquilo que o teste mede e o quão bem ele faz isso” (Anastasi & Urbina, 2000, p. 107). Para Cohen, Swerdlik e Sturman (2014) a validade de um teste é “um julgamento ou uma estimativa de quão bem ele mede o que se propõe a medir em um determinado contexto” (p. 181). Segundo estes autores, a validade constitui um julgamento baseado nas evidências

sobre a adequação de inferências que podem ser extraídas a partir das pontuações dos testes.

A Validade do Koppitz-2 foi investigada de diferentes formas: 1) pela diferenciação entre idades, também chamada de validade de desenvolvimento; 2) pela comparação com outros testes, que pode ser considerada validade de critério concorrente, ou validade de construto e 3) por meio da relação entre o sistema Koppitz-2 (Reynolds, 2007) e o original (Koppitz, 1963/1989). Para o segundo estudo, foram calculadas as correlações com o Teste de Inteligência Não Verbal – R-2 e com o Teste das Figuras Complexas de Rey, os resultados destes últimos são apresentados a seguir.

2.1) Validade com o Teste das Figuras Complexas de Rey

A validade do Koppitz-2 foi verificada pela correlação com o Teste das Figuras Complexas de Rey. Trata-se de estudo de validade em que se busca verificar a relação entre os resultados de dois instrumentos que avaliam o mesmo construto, ou seja, a percepção visual e a capacidade de integração visomotora. Segundo Cohen et al. (2014, p. 202):

Se os escores no teste que está sendo submetido à validação de construto tenderem a forte relação na direção prevista com escores em testes mais antigos, mais estabelecidos e já validados que visam medir o mesmo construto (ou um construto semelhante), isso seria um exemplo de evidência convergente.

Conforme mencionado anteriormente, as crianças de 6 e 7 anos foram avaliadas pela Figura B do Teste das Figuras Complexas de Rey e as de 10 e 11 anos pela Figura A. Os resultados serão apresentados separadamente para os dois grupos.

Para 6 e 7 anos a média do tempo de aplicação do Koppitz-2 foi de 7,43 minutos e o desvio padrão de 2,52. Para as Figuras Complexas de Rey a média de tempo foi de 2,50 minutos e o desvio padrão 1,24. Os resultados das correlações entre os totais de pontos do Koppitz-2 e da Figura B para a amostra total e para os grupos em função da idade e do sexo, são apresentados na Tabela 36.

Tabela 36. Correlações (*r*) entre os pontos da Figura de Rey B e Koppitz-2 para os mais novos por idade e por sexo

Idades	<i>r</i>	N
6	0,616**	17
7	0,648**	32
Sexo		
Feminino	0,596**	24
Masculino	0,690**	25
Total	0,639**	49

** Significantes a 0,01

A correlação entre o total de pontos da Figura B e Koppitz-2 para o grupo de 6 e 7 anos foi de 0,639 (estatisticamente significativa a 0,01), considerada de magnitude alta, segundo os critérios de Sisto (2007) e moderada de acordo com Dancey e Reidy (2013), o que indica uma relação moderada entre os resultados dos dois testes. A correlação para 7 anos foi maior do que a de 6 (respectivamente 0,648 e 0,616), ambas estatisticamente significantes. No entanto, a correlação para 6 anos deve ser considerada com atenção já que o número de crianças avaliadas nesta idade foi menor. Quanto às correlações por sexo, pode-se constatar que a dos meninos foi maior do que a das meninas (respectivamente 0,690 e 0,596).

É importante ressaltar que as médias desta pesquisa foram menores do que as do manual do Rey. Para 6 anos, a média do manual é 21,72 pontos e desvio padrão 6,11, enquanto nesta pesquisa a média foi de 15,18 e o desvio padrão de 5,74 ($t=2,006$; $p<0,05$). Para 7 anos, a média do manual é 24,81 pontos e o desvio padrão 4,59 e nesta, a média foi de 20,87 e o desvio padrão 4,18 ($t=4,237$; $p<0,05$). Deve ser lembrado ainda, que no manual do Teste das Figuras Complexas de Rey para a Figura B, não houve discriminação quanto ao tipo de escola, bem como de que a amostra desta pesquisa é menor do que a do manual.

A Tabela 37 traz os resultados das correlações entre o Koppitz-2 e a Figura A das Figuras Complexas de Rey para a amostra total e para os grupos em função da idade e do sexo para os participantes com 10 e 11 anos.

Tabela 37. Correlações (r) entre a Figura A e Koppitz-2 para os mais velhos por idade e sexo

Idades	r	
10	0,647**	26
11	0,778**	25
Sexo		
Feminino	0,759**	28
Masculino	0,637**	23
Total	0,720**	51

** Significantes a 0,01

A correlação entre o total de pontos da Figura A e o Koppitz-2 de 10 e 11 anos foi alta (0,720, estatisticamente significativa a 0,01). Para 11 anos a correlação foi maior do que para 10 anos (respectivamente 0,778 e 0,647) e a das meninas maior do que a dos meninos (0,759 e 0,637 respectivamente). Todas as correlações também foram estatisticamente significantes e com magnitude de moderada a alta de acordo com os parâmetros de Sisto (2007) e Dancey e Reidy (2013).

Comparando as correlações dos mais jovens (6 e 7 anos) com a dos mais velhos (10 e 11 anos) as destes últimos foram mais elevadas para o grupo total. As correlações de cada grupo etário foram maiores para os 11 anos, seguidas pelas de 7 anos. Esses resultados constituem evidência da validade convergente entre os Testes de Rey e Koppitz-2, uma vez que as correlações encontradas estão dentro dos parâmetros considerados adequados pela literatura científica (Guilford, 1950; Nick & Kellner, 1971, Fonseca, 2003), o que permite afirmar que há uma alta relação entre os dois instrumentos indicando que avaliam a percepção visual e a integração visomotora.

Segundo Nunes e Primi (2010, p. 118),

para que o instrumento cujo manual está sendo avaliado seja reconhecido como tendo evidência de validade convergente, ele deve conter no mínimo um estudo que indique correlação compatível com esse tipo de validade. Se dois testes medem construtos idênticos, devem ter altas correlações entre si. Portanto, os coeficientes devem ser de magnitude alta, usualmente acima de 0,509, para se considerar como um instrumento que avalie o mesmo construto, e que esse outro teste tenha evidências reconhecidas de validade e fidedignidade no Brasil.

As médias das crianças desta pesquisa para 10 e 11 anos foram menores do que as do manual do Rey. Para 10 anos, a média do manual é 29 pontos e desvio padrão 6,27, nesta pesquisa a média foi de 21,09 pontos e o desvio padrão de 5,96 ($t=5,080$; $p<0,01$). Para 11 anos, a média do manual é 31,77 e o desvio padrão 3,29, nesta pesquisa a média foi de 23,58 pontos e o desvio padrão de 6,58 ($t=5,728$; $p<0,01$). No entanto, é importante destacar, que o manual do Teste das Figuras Complexas de Rey para a Figura A não diferencia as crianças pelo tipo de escola.

2.2) Validade com o Teste R-2

A validade também foi obtida pela correlação entre o total de pontos do R-2 e do Koppitz-2.

A média do tempo de execução do Koppitz-2 foi de 9,81 minutos e o desvio padrão de 4,92. Para o R-2 a média de aplicação foi de 6,88 minutos e o desvio padrão 2,29. As correlações entre os totais de pontos do R-2 e do Koppitz-2 para a amostra total e para os grupos em função da idade e do sexo dos participantes, podem ser observadas na Tabela 38.

Tabela 38. Correlações (r) entre o R-2 e Koppitz-2 para os mais novos

Idades	r	N
6	0,638**	21
7	0,523**	31
Sexo		
Feminino	0,542**	27
Masculino	0,620**	25
Total	0,594**	52

*** Significantes a 0,01

A correlação entre o total de pontos do R-2 e Koppitz-2 para 6 e 7 anos foi de 0,594 (estatisticamente significativa a 0,01). Sisto (2007) e Dancey e Reidy (2013) consideram moderadas as correlações entre 0,40 e 0,60, indicando que há uma relação adequada entre os dois testes. Pode-se verificar ainda que a correlação dos participantes com 6 anos foi maior do que do grupo de 7 anos (respectivamente 0,638 e 0,523) e dos meninos maior do que as meninas (0,620 para o primeiro e

para o último 0,542), sendo todas estatisticamente significantes e de magnitude moderada.

As médias das crianças desta pesquisa foram maiores do que a da pesquisa do manual do R-2 em relação às crianças de escolas públicas por idade. Para 6 anos, a média do manual é 9,5 pontos e desvio padrão 3,7, enquanto nesta pesquisa a média foi de 12,62 e o desvio padrão de 4,37 ($t=3,068$; $p<0,01$). Para 7 anos, a média do manual é 10,9 pontos e o desvio padrão 4,2, nesta pesquisa a média foi de 15,9 e o desvio padrão 5,36 ($t=4,878$; $p<0,01$).

Os resultados das correlações entre os totais de pontos do R-2 e do Koppitz-2 para a amostra total e para os grupos em função da idade e do sexo com 10 e 11 anos, podem ser observados na Tabela 39.

Tabela 39. Correlações entre o R-2 e Koppitz-2 para os participantes mais velhos

Idades	<i>r</i>	N
10	0,439**	37
11	0,667**	16
Sexo		
Feminino	0,520**	26
Masculino	0,568**	27
Total	0,539**	53

* Significantes a 0,05

** Significantes a 0,01

A correlação entre o total de pontos do R-2 e Koppitz-2 para o grupo de 10 e 11 anos foi de 0,539 (significante a 0,01). Pode-se constatar ainda, que a correlação de 11 anos foi maior do que a de 10 (respectivamente 0,667 e 0,439) e dos meninos maior do que as meninas (respectivamente 0,568 e para 0,520), sendo todas estatisticamente significantes. Sisto (2007) e Dancey e Reidy (2013) consideram moderadas as correlações entre 0,40 e 0,60, indicando uma relação moderada entre os dois testes. Cabe destacar que a correlação com os participantes de 11 anos deve ser considerada com atenção uma vez que o grupo foi reduzido (N=16).

Quando se compara os mais velhos com os dos mais novos, verifica-se que as correlações deste último grupo são um pouco maiores para a amostra total. Os resultados por faixa etária, mostram que as correlações de 11 anos foram maiores, seguidas das de 7 anos.

Estes resultados confirmam a validade simultânea entre o R-2 e o Koppitz-2 tanto para os participantes de 6 e 7 quanto para os de 10 e 11 anos, pois para Guilford (1950) um coeficiente de validade pode variar entre 0,0 e 0,60 “*com muitos índices na metade inferior desta faixa*” (p.165) e completa afirmando que “*qualquer coeficiente de correlação que não é zero e é estatisticamente significativa indica algum grau de relação entre as variáveis*” (p.164). Este autor apresenta uma tabela de interpretação de correlação que considera como moderadas as correlações entre 0,40 e 0,70 e sugerem uma relação substancial entre as variáveis estudadas (Fonseca, 2003). Nick e Kellner (1971) propõem que as correlações adequadas para dois testes diferentes que avaliam a mesma característica devem oscilar entre 0,60 e 0,70. Pode-se dizer, portanto, que as correlações obtidas com o R-2 estão de acordo com os parâmetros de Guilford (1950) e muito próximos dos apresentados por Nick e Kellner (1971), caracterizando-se como um coeficiente de validade adequado. Estes resultados indicam a existência de uma relação entre os dois instrumentos, isto é, entre a integração visomotora, que está relacionada à capacidade de integrar estímulos visuais e respostas motoras, e a capacidade cognitiva. A partir disso pode-se inferir que a integração visomotora tenha um papel importante no desenvolvimento cognitivo o que aponta para a possibilidade de utilização do Koppitz-2 para a avaliação de crianças com dificuldades de aprendizagem.

A média das crianças desta pesquisa para 10 anos foi maior do que a do manual do R-2, em relação às crianças de escolas públicas. Para 10 anos, a média do manual é 17,8 pontos e desvio padrão 5,3, nesta pesquisa a média foi de 20,81 e o desvio padrão de 4,03 ($t=3,889$; $p<0,01$). Para 11 anos, ainda que a média da pesquisa tenha sido maior, a diferença não foi estatisticamente significativa (a média do manual é 18,9 pontos e o desvio padrão 5,0, nesta pesquisa a média foi de 19,94 pontos e o desvio padrão 5,18 ($t=0,753$; $p>0,05$). No entanto, é importante observar que o grupo de 11 anos nesta amostra foi bastante reduzido o que pode ter influenciado nos resultados.

2.3) Relação entre os sistemas Koppitz-2 e o original

Conforme mencionado anteriormente foram selecionados 300 protocolos da amostra total que foram avaliados também pelos indicadores Maturacionais de Koppitz (1963/1989). Os totais de pontos dos dois sistemas foram correlacionados,

considerando os resultados das crianças de 6 e 7 e de 8 a 12 anos separadamente. Os resultados são apresentados a seguir.

Tabela 40. Correlações entre os sistemas Koppitz-2 e Koppitz (1963/1989) por idade

Idades	Todos	9 Figuras
6	-0,628**	-0,633**
7	-0,601**	-0,600**
8	-0,511**	-0,602**
9	-0,501**	-0,431**
10	-0,441**	-0,482**
11	-0,717**	-0,717**
12	-0,639**	-0,654**

** Significantes a 0,01

Foram realizadas análises, tanto para o total de pontos de todos os modelos relativos à idade do examinando do Koppitz-2, quanto para apenas os nove do sistema original. Para todos os modelos, as correlações variaram entre -0,441 e -0,717, consideradas de magnitude moderada a alta (Sisto, 2007 e Dancey & Reidy, 2013). A correlação mais alta foi para a idade de 11 anos (-0,717) e a mais baixa de 10 anos (-0,441). As correlações foram negativas, pois no sistema original pontua-se “os erros”, enquanto no Koppitz-2 são dados pontos para os acertos, ou seja, para a qualidade da reprodução.

As correlações com as nove figuras comuns aos dois sistemas, oscilaram entre -0,431 e -0,717, também neste caso, foram mais elevadas aos 11 e mais baixas aos 9 anos. Pode-se observar a tendência de um aumento das correlações somente para o total de nove figuras, o que não ocorreu apenas para 9 anos (para 7 anos a correlação também foi menor, mas muito próxima à de todos os modelos).

Tabela 41. Correlações entre os sistemas Koppitz-2 e Koppitz (1963/1989) por faixa etária e sexo para todas as figuras por idade ou apenas para as nove comuns

Idades	Feminino		Masculino		Total	
	Todas	Nove	Todas	Nove	Todas	Nove
6 e 7	-0,476**	-0,471**	-0,791**	-0,789**	-0,647**	-0,646**
8 a 12	-0,566**	-0,575**	-0,566**	-0,580**	-0,557**	-0,566**

** Significantes a 0,01

As correlações para 6 e 7 anos variaram entre -0,471 e -0,789 e para 8 a 12 entre -0,557 e -0,580. Os resultados indicam uma relação entre os dois sistemas de avaliação e, em geral, superiores ao parâmetro apontado por Nunes e Primi (2010) para testes que avaliam construtos relacionados o que indica evidência de validade para o Koppitz-2.

3) Fidedignidade

A fidedignidade de um teste mostra o quanto seus resultados são confiáveis (Urbina, 2007) e se refere à segurança da mensuração (Cohen et al., 2014). Pode ainda ser definida como “a extensão em que as diferenças individuais nos escores dos testes são atribuíveis a diferenças ‘verdadeiras’ nas características sob consideração e a extensão em que elas são atribuíveis a erros casuais” (Anastasi & Urbina, 2000, p. 84).

A Fidedignidade do Koppitz-2 foi avaliada por três métodos: Teste e Reteste, Precisão entre Avaliadores e Consistência Interna. Os resultados são apresentados a seguir.

3.1) Método do Teste-Reteste

O Teste-reteste é “uma estimativa de confiabilidade obtida pela correlação de pares de escores da mesma pessoa em duas administrações diferentes do mesmo teste” (Cohen et al., 2014, p. 150). O coeficiente de precisão obtido por este método também constitui uma medida de estabilidade do teste, já que cada examinando é avaliado duas vezes pelo mesmo instrumento em ocasiões diferentes, separadas por um intervalo de tempo (Urbina, 2007).

A precisão pelo método do teste-reteste foi obtida por meio da correlação entre os totais de pontos das crianças em duas aplicações que ocorreram com um intervalo médio entre as aplicações de aproximadamente 35,7 dias (variando entre 27 e 54 dias). Foram calculados para as crianças entre 6 e 7 anos e entre 10 e 11 anos porque os desenhos e critérios de avaliação do Koppitz-2 são diferentes para as duas faixas etárias. Também foram comparadas as médias do total de pontos nas

duas aplicações por meio do *Teste t de Student* para amostras pareadas para investigar o efeito de aprendizagem.

Para verificar, a existência de diferença estatisticamente significativa entre o tempo utilizado pelas crianças de 6 e 7 anos para a realização dos desenhos no teste e no reteste foram comparadas as médias por meio do *Teste t de Student*. A média do tempo para o teste foi de 8,20 minutos (DP=2,79) e para o reteste de 7,08 (DP=2,20). O resultado mostrou que a diferença entre as médias foi estatisticamente significativa ($t=2,502$; $p<0,05$) e indica que as crianças tenderam a copiar com maior rapidez no reteste do que no teste.

A Tabela 42 apresenta os resultados das correlações entre os totais de pontos do teste e do reteste para a amostra total de mais novos e para os grupos em função da idade e do sexo.

Tabela 42. Correlações (r) para amostra total por idade e sexo dos mais novos – grupo do reteste

Idades	r	N
6	0,837**	23
7	0,750**	28
Sexo		
Feminino	0,750**	26
Masculino	0,822**	25
Total	0,794**	51

** Significantes a 0,01

A correlação entre o total de pontos do teste e do reteste para o grupo de 6 e 7 anos foi de 0,794 (estatisticamente significativa a 0,01). Guilford (1950) considera adequados valores entre 0,70 e 0,98 para coeficientes de precisão. Segundo Anastasi e Urbina (2000) são esperados coeficientes de precisão entre 0,80 e 0,90, dessa forma pode-se dizer que o resultado observado neste estudo encontra-se muito próximo destes parâmetros o que é uma evidência de fidedignidade para o Koppitz-2 no grupo de 6 e 7 anos. A Tabela 42 mostra ainda que a correlação de 6 anos foi maior do que a de 7 anos (respectivamente 0,837 e 0,750) e a dos meninos maior do que a das meninas (0,822 e 0,750). Em todos os casos os coeficientes foram altos e indicaram forte relação entre os resultados do teste e do reteste (Sisto, 2007; Dancey & Reidy, 2013).

A comparação entre as médias nos dois momentos de testagem, quando se considera a faixa de 6 e 7 anos foi estatisticamente significativa ($t=-3,091$; $p<0,01$). A média do total de pontos no teste foi de 15,45 pontos (DP=4,11) e no reteste 16,57 pontos (DP=3,91). Pode-se verificar, portanto, que a média do reteste foi um pouco maior do que a do teste, indicando a ocorrência de um efeito de aprendizagem na execução dos desenhos.

Quanto à idade, as diferenças entre as médias não foram estatisticamente significantes para as crianças de 6 anos (média de 14,83 no teste, DP=4,71 e no reteste 15,17 e DP=4,17; $t=-0,646$, $p>0,01$), mas foram para as de 7 anos (média de 15,96 pontos no teste, DP=3,55 e para o reteste 17,71 pontos, DP=3,33; $t=-3,789$, $p<0,01$). Em relação à variável sexo, as diferenças foram estatisticamente significantes, tanto para as meninas (média de 16,15 pontos no teste, DP=3,55 e para o reteste 17,31 pontos, DP=3,78; $t=-2,260$, $p<0,05$), quanto para os meninos (média de 14,72 pontos no teste, DP=4,58 e para o reteste 15,80 pontos, DP=3,97; $t=-2,067$, $p<0,05$).

A existência de uma diferença estatisticamente significativa entre o tempo utilizado para a realização dos desenhos no teste e no reteste foi verificada através do *Teste t de Student*. A média de tempo para o teste foi de 11,09 minutos (DP=3,73) e para o reteste de 9,89 (DP=4,04). A diferença entre as médias foi estatisticamente significativa ($t=2,174$; $p<0,05$) e indica que os participantes tenderam a realizar os desenhos com maior rapidez no reteste.

As correlações entre os totais de pontos do teste e do reteste para a amostra de 10 e 11 anos e para os grupos em função da idade e do sexo podem ser observadas na Tabela 43.

Tabela 43. Correlações (r) para amostra de 10 e 11 anos por idade e sexo – grupo do reteste

Idades	r	N
10	0,745**	38
11	0,964**	12
Sexo		
Feminino	0,911**	26
Masculino	0,769**	27
Total	0,850**	50

** Significantes a 0,01

A correlação entre o total de pontos entre o teste e o reteste para 10 e 11 anos foi de 0,850 (significante a 0,01). Segundo Sisto (2007), este valor é considerado muito alto. Dancey e Reidy (2013) afirmam que as correlações entre 0,70 e 0,90 são fortes e significam uma forte relação entre os totais das duas aplicações. Anastasi e Urbina (2000) afirmam que em estudos de teste e reteste são esperados coeficientes de precisão entre 0,80 e 0,90. Pode-se dizer, portanto, que o resultado deste estudo encontra-se dentro dos parâmetros aceitos na literatura científica, o que mostra evidência de fidedignidade para o Koppitz-2 nesta faixa etária. A Tabela 43 permite verificar também que a correlação de 11 anos foi maior do que a de 10 anos (respectivamente 0,964 e 0,745) e das meninas maior do que a dos meninos (0,911 e 0,769). No entanto, os resultados de 11 anos devem ser considerados com cautela, uma vez que o grupo foi bastante reduzido na amostra estudada. Todas as correlações foram altas e indicaram forte relação entre os resultados das duas aplicações (Sisto, 2007; Dancey & Reidy, 2013).

A comparação entre as médias das duas aplicações para a faixa de 10 e 11 anos foi estatisticamente significativa ($t=-2,301$; $p<0,05$). A média do total de pontos no teste foi de 20,68 pontos (DP=6,59) e no reteste 21,84 pontos (DP=6,39). Pode-se verificar, portanto, que a média do reteste foi um pouco maior do que a do teste, o que significa que na amostra estudada ocorreu um efeito de aprendizagem na forma como os examinandos responderam ao teste.

Quanto à idade, as diferenças de médias foram estatisticamente significantes para as crianças de 10 anos (média de 20,50 no teste, DP=5,61 e no reteste 21,76 pontos, DP=5,05; $t=-2,028$, $p<0,05$) e não para 11 anos (média de 21,25 no teste, DP=9,36 e no reteste 22,08, DP=9,81; $t=-1,101$, $p>0,05$). Cabe ressaltar novamente que estes resultados devem ser considerados com cautela, pois o grupo de 11 anos era bastante reduzido neste estudo. Em relação ao sexo, tanto para as meninas (média de 21,96 no teste, DP=7,50 e 22,88, DP=6,44 no reteste, $t=-1,507$, $p>0,05$), quanto para os meninos (média de 19,29 no teste e DP=5,27 e para o reteste, média de 20,71 e DP=6,28; $t=-1,717$, $p>0,05$) os resultados não foram estatisticamente significantes.

3.2) Consistência interna

A fidedignidade de Koppitz-2 foi obtida pelo cálculo do Alfa de Cronbach. Segundo Pasquali (2003) o Alfa de Cronbach é uma técnica de “*estimativa de coeficientes de precisão*” que tem como objetivo verificar “*a consistência interna do teste através da análise da consistência interna dos itens*”, ou seja, permite constatar a “*congruência que cada item do teste tem com o restante dos itens do mesmo teste*” (p.203). A Tabela 44 mostra os resultados para as crianças de 6 e 7 anos, por idade e sexo.

Tabela 44. Alfa de Cronbach para o Koppitz-2 para 6 e 7 anos por idade e sexo

Idades	Alfa	N
6	0,810	68
7	0,716	85
Sexo		
Feminino	0,785	76
Masculino	0,793	77
Total	0,794	153

Segundo Hair, Black, Babin, Anderson e Tatham (2009) os alfas de 0,60 a 0,70 são considerados como limítrofes para que um resultado seja aceitável. A Tabela 44 mostra que quando se considera toda a faixa etária o alfa foi de 0,794, os alfas dos meninos (0,793) foram maiores do que o das meninas (0,785) e os de 6 anos (0,810) maiores do que os de 7 anos (0,716). A Tabela 45 apresenta os alfas entre 8 e 12 anos por idade, sexo e tipo de escola.

Tabela 45. Alfa de Cronbach para o Koppitz-2 por idade, sexo e tipo de escola para 8 a 12 anos

Idades	Alfa	N
8	0,819	102
9	0,847	91
10	0,815	119
11	0,817	87
12	0,879	71
Sexo		
Feminino	0,842	246
Masculino	0,842	224

Tipo de escola		
Estadual	0,832	302
Municipal	0,858	168
Total	0,843	470

É possível verificar pela Tabela 45 que o Alfa de Cronbach para toda a faixa etária foi de 0,843. Pode-se constatar ainda que quanto à idade, ao sexo e ao tipo de escola em todos os casos os coeficientes foram superiores a 0,80 indicando um índice muito bom de precisão para o Koppitz-2.

Portanto é possível afirmar que nos dois grupos (6 e 7 anos e 8 a 12 anos) os resultados foram superiores à faixa apontada por Hair et al. (2009) o que significa que há uma adequada consistência interna entre os itens do Koppitz-2, indicando fidedignidade satisfatória para o sistema de avaliação.

3.3) Precisão entre avaliadores

Como já foi mencionado anteriormente, para o estudos de precisão entre avaliadores foram selecionados 62 protocolos, sendo 31 de crianças entre 6 e 7 anos e 31 de 8 a 12 anos, de ambos os sexos. Os participantes foram selecionados aleatoriamente, mas de forma a abranger todas as faixas de idade para evitar qualquer influência desta variável no estudo. Os protocolos foram submetidos a avaliações de dois psicólogos especialistas no Bender que foram treinados na avaliação Koppitz-2 (foram realizadas supervisões das avaliações para garantir maior compreensão dos critérios que seriam aplicados nas análises).

Verificou-se que as correlações entre os totais de pontos dos dois avaliadores para o Sistema Koppitz-2 foram de 0,966 para as crianças de 6 e 7 anos e 0,972 para as entre 8 e 12 anos, ambas consideradas muito altas e estatisticamente significantes (Sisto, 2007 e Dancey & Reidy, 2013). Os resultados indicaram coeficientes de concordância entre avaliadores elevados e superiores a 0,70, constituindo evidência de fidedignidade do instrumento (Anastasi & Urbina, 2000).

As médias das pontuações dos juízes foram comparadas por meio do *Teste t de Student*. As diferenças não foram estatisticamente significantes tanto quando se considera o grupo de mais novos (média do Juiz 1= 15,90 e DP=5,67 e para o Juiz 2

média 15,77 e DP=5,45; $t=0,486$, $p>0,05$) quanto os mais velhos (média do Juiz 1= 25,84 e DP=7,58 e para o Juiz 2 média 25,61 e DP=7,51; $t=0,698$, $p>0,05$).

Para verificar a concordância entre os avaliadores para os itens do Koppitz-2 foram calculados os Coeficientes Kappa de Concordância. A Tabela 46 apresenta os resultados para cada item do sistema Koppitz-2 para as crianças de 6 e 7 anos.

Tabela 46. Índices Kappa para os itens do Koppitz-2 para crianças entre 6 e 7 anos

Itens	Juiz 1 x Juiz 2	Itens	Juiz 1 x Juiz 2
1	1,000**	18	0,839**
2	0,475**	19	0,679**
3	0,870**	20	0,793**
4	0,870**	21	0,859**
5	0,737**	22	-
6	0,652**	23	0,870**
7	0,784**	24	0,868**
8	0,903**	25	0,762**
9	0,919**	26	0,817**
10	0,739**	27	0,631**
11	0,763**	28	0,871**
12	0,665**	29	0,832**
13	0,807**	30	0,784**
14	0,817**	31	0,762**
15	0,805**	32	0,451**
16	0,839**	33	0,674**
17	0,935**	34	0,528**

** Significantes a 0,01

A Tabela 46 mostra que dos 34 itens, 17 (50,0%) apresentaram concordância entre 0,80 e 1,00, considerada quase perfeita; 13 (38,2%) substancial (entre 0,60 e 0,79) e 3 (8,8%) moderada (entre 0,40 e 0,59), segundo os parâmetros de Baltar e Okano (s.d), Byrt (1996) e Landis e Koch (1977). Para o item 22, não foi possível realizar o cálculo, pois uma das variáveis foi constante. No entanto, a análise da frequência das pontuações mostrou que os juízes discordaram em apenas dois casos, o que indica um alto grau de concordância. Na Tabela 47 estão os resultados do Kappa para cada item do sistema Koppitz-2 para as crianças de 8 e 12 anos.

Tabela 47. Índices Kappa para os itens do Koppitz-2 crianças entre 8 e 12 anos

Itens	Juiz 1 x Juiz 2	Itens	Juiz 1 x Juiz 2
1	-	24	0,839**
2	0,652**	25	0,652**
3	0,773**	26	0,795**
4	-	27	0,705**
5	0,760**	28	0,563**
6	0,680**	29	0,554**
7	0,600**	30	0,668**
8	0,570**	31	0,933**
9	0,658**	32	0,865**
10	0,735**	33	0,652**
11	0,652**	34	0,688**
12	0,871**	35	0,773**
13	0,676**	36	0,729**
14	0,674**	37	0,757**
15	0,853**	38	0,669**
16	0,762**	39	1,000**
17	0,713**	40	0,865**
18	1,000**	41	0,817**
19	0,919**	42	0,933**
20	0,868**	43	0,871**
21	0,919**	44	0,652**
22	0,563**	45	0,868**
23	0,853**		

** Significantes a 0,01

A Tabela 47 mostra que dos 45 itens, 16 (35,6%) indicaram concordância quase perfeita, 23 (51,2%) concordância substancial e 4 (8,8%) moderada, sendo que para dois itens (4,4%), ele não foi calculado. O Kappa para os itens 1 e 4 não puderam ser calculados pois não ocorreu variabilidade de pontuação na avaliação de um dos juízes. A análise da frequência das pontuações nestes dois itens mostrou que os juízes concordaram em 93,5% dos casos (em apenas dois protocolos as avaliações foram diferentes) o que representa um adequado índice de concordância.

Uma das fontes de erros na mensuração/avaliação dos testes é originária das avaliações realizadas pelos psicólogos e profissionais (Reynolds, 2007). A precisão entre avaliadores tem a finalidade de verificar a ocorrência da variação entre os julgamentos dos avaliadores, o que pode interferir nos resultados do teste. As correlações deste trabalho indicaram um adequado grau de concordância entre as

avaliações dos dois juízes, tanto para as crianças de mais novas, quanto para as mais velhas, uma vez que todos os itens tiveram coeficientes, mostrando uma concordância, que variou entre moderada e quase perfeita. As correlações entre os totais de pontos também foram muito altas (Sisto, 2007, Reidy & Dancey, 2013) confirmando a fidedignidade do instrumento por este método, principalmente considerando que a avaliação é feita em função do total de pontos do protocolo e não dos itens isoladamente.

4) Análise de Itens

A análise dos itens do Koppitz-2 foi realizada pelo cálculo da correlação item-total e pela comparação das frequências de pontos nos itens do teste em função da idade, por meio do cálculo do *Qui-Quadrado* para determinar diferenças nos coeficientes em relação à idade.

4.1) Resultados para as crianças de 6 e 7 anos

A Tabela 48 mostra as correlações item-total para cada item por sexo e idade e para a faixa de 6 e 7 anos.

Tabela 48. Correlação Item-Total (r_{it}) por sexo, idade e para amostra total para as idades de 6 e 7 anos

Itens	Sexo		Idade		Total
	Feminino	Masculino	6	7	
1	0,439	0,346	0,417	0,160	0,392
2	0,161	0,129	0,088	0,152	0,120
3	0,227	0,349	0,374	0,204	0,317
4	0,359	0,265	0,333	0,226	0,312
5	0,193	0,334	0,219	0,364	0,269
6	0,157	0,556	0,495	0,086	0,422
7	0,347	0,349	0,476	0,116	0,365
8	0,503	0,461	0,475	0,376	0,486
9	0,483	0,495	0,572	0,274	0,495
10	0,275	0,200	0,290	0,157	0,251
11	0,326	0,283	0,282	0,250	0,296
12	-0,137	0,177	0,018	0,039	0,022
13	0,135	0,169	0,200	0,062	0,149
14	0,290	0,303	0,276	0,335	0,282
15	0,511	0,479	0,450	0,505	0,503
16	0,318	0,240	0,029	0,326	0,283
17	0,284	0,323	0,274	0,253	0,307
18	0,065	0,277	0,179	0,142	0,182
19	0,240	0,391	0,371	0,159	0,319
20	0,462	0,161	0,310	0,282	0,307
21	0,244	0,265	0,232	0,292	0,267
22	0,152	0,000	0,076	0,200	0,122
23	0,209	0,112	0,224	0,117	0,175
24	0,255	0,044	0,164	0,156	0,163
25	0,177	-0,018	0,276	-0,054	0,071
26	0,219	0,276	0,220	0,227	0,252
27	0,344	0,283	0,194	0,356	0,323
28	0,261	0,414	0,480	0,140	0,335
29	0,444	0,481	0,489	0,356	0,463
30	0,329	0,350	0,369	0,245	0,340
31	0,431	0,088	0,150	0,417	0,267
32	0,352	0,392	0,461	0,268	0,365
33	0,200	0,181	0,192	0,097	0,173
34	0,396	0,414	0,257	0,450	0,366

Guilford e Frutcher (1978) consideram que o valor mínimo da correlação item-total para que um item seja discriminativo é de 0,30. Rueda e Muniz (2012) e

Rueda e Castro (2012) afirmam que o *American Institute for Research* considera o valor de 0,20 como indicador de boa discriminação do item. Considerando este parâmetro, pode-se verificar que dos 34 itens que compõem o teste, nove (26,4%) tiveram resultados inferiores. Os itens foram: 2 (Figura 2); 12 e 13 (Figura 7); 18 (Figura 9); 22 (Figura 10); 23, 24, 25 (Figura 11) e 33 (Figura 13). Para verificar, se a exclusão dos itens, que tiveram correlação item-total baixa alterava significativamente o valor do Alfa de Cronbach, foram refeitos os cálculos seguindo este critério. Apesar do valor ter sido um pouco mais elevado (0,809) a alteração não foi significativa. Também foi verificado, se a exclusão de cada item isoladamente alteraria o alfa total e os resultados estão na Tabela 49:

Tabela 49. Variações dos alfas com a exclusão de itens com baixa correlação item-total - grupo de 6 e 7 anos

Itens	Alfa com item excluído
2	0,794
12	0,798
13	0,796
18	0,793
22	0,794
23	0,794
24	0,795
25	0,796
33	0,793

Assim como em relação à retirada de todos os itens com baixa correlação item-total, quando isso é feito isoladamente, pode-se constatar que os valores não sofreram grandes alterações.

A Tabela 50 traz a frequência de pontuação dos itens do Koppitz-2 das crianças entre 6 e 7 anos. Para verificar se os grupos se diferenciavam em função da idade foi calculado o *Qui-quadrado* para cada um dos itens.

Tabela 50. Frequência percentual de acerto dos itens e qui-quadrado por idade para amostra de 6 e 7 anos

Itens	Idade		Total	Qui-Quadrado	Sig
	6	7			
1	79,4	96,6	89,5	13,328**	0,000
2	94,4	95,3	94,8	0,105	0,511
3	70,6	81,2	76,5	2,354	0,125
4	86,8	94,1	90,8	2,457	0,117
5	35,3	35,3	35,3	0,000	1,000
6	83,8	98,8	92,2	11,760**	0,001
7	54,4	77,6	67,3	9,271**	0,002
8	63,2	87,1	76,5	11,916**	0,001
9	61,8	85,9	75,2	11,771**	0,001
10	25,0	38,8	32,7	3,281	0,070
11	70,6	82,4	77,1	2,964	0,085
12	16,2	15,3	15,7	0,022	0,881
13	35,3	43,5	39,9	1,069	0,301
14	25,0	24,7	24,8	0,002	0,967
15	30,9	50,6	41,8	6,029*	0,014
16	1,5	17,6	10,5	10,557**	0,001
17	39,7	58,8	50,3	5,523*	0,019
18	20,6	29,4	25,5	1,549	0,213
19	54,4	74,1	65,4	6,479*	0,011
20	54,4	63,5	59,5	1,303	0,254
21	54,4	61,2	58,2	0,710	0,399
22	1,5	1,2	1,3	0,025	0,874
23	52,9	57,6	55,6	0,339	0,561
24	44,1	48,2	46,4	0,258	0,612
25	19,1	9,4	13,7	3,005	0,083
26	10,3	21,2	16,3	3,273	0,070
27	1,5	14,1	8,5	7,772**	0,005
28	44,1	58,8	52,3	3,275	0,070
29	58,8	77,6	69,3	6,290*	0,012
30	47,1	63,5	56,2	4,164*	0,041
31	8,8	9,4	9,2	0,016	0,900
32	17,6	27,1	22,9	1,897	0,168
33	7,4	17,6	13,1	3,253	0,061
34	5,9	11,8	9,2	1,572	0,210

** Significante a 0,01

* Significante a 0,05

A Tabela 50 mostra que para todos os itens, a frequência de pontuação foi maior para sete anos do que para seis. Apesar disso, somente em 12 itens as

diferenças foram estatisticamente significantes, com frequências menores de pontuação para o grupo mais novo. Os itens foram 1 (Figura 1); 6 (Figura 4); 7, 8 e 9 (Figura 5); 15 e 16 (Figura 8); 17 e 19 (Figura 9); 27, 29 e 30 (Figura 12).

A Figura 6 mostra a frequência percentual de cada item para 6 e 7 anos e para as duas idades reunidas.

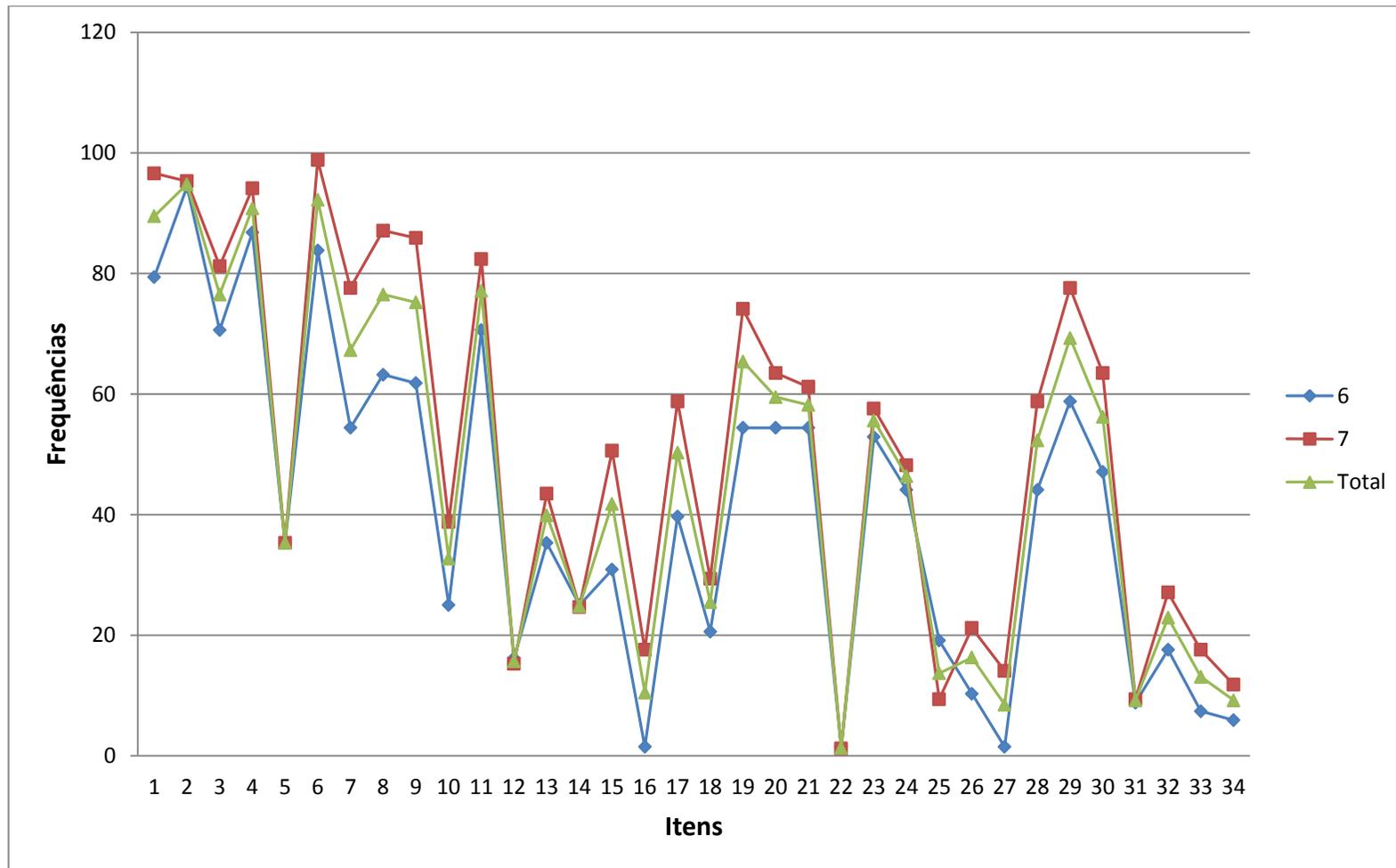


Figura 6. Frequência percentual de pontos dos itens por idade e para a amostra total de 6 e 7 anos

4.2) Resultados para as crianças entre 8 e 12 anos

A Tabela 51 mostra a correlação item-total de cada um dos itens do Koppitz-2 por idade e para a amostra total de 8 a 12 anos.

Tabela 51. Correlação Item Total (r_{it}) por idade e para amostra total de 8 a 12 anos

Itens	Idade					Total
	8	9	10	11	12	
1	0,193	0,194	0,282	0,236	0,147	0,215
2	0,350	0,209	0,113	0,198	0,032	0,189
3	0,232	0,295	0,167	0,166	0,339	0,231
4	0,257	0,060	0,133	0,119	0,120	0,156
5	0,229	0,156	0,122	-0,102	0,255	0,126
6	0,155	0,367	0,086	0,411	0,380	0,261
7	0,011	-0,044	0,124	0,123	0,047	0,068
8	0,121	0,072	0,138	0,082	0,125	0,132
9	0,181	0,301	0,191	0,208	0,415	0,267
10	0,450	0,374	0,441	0,214	0,430	0,390
11	0,307	0,128	0,261	0,434	0,157	0,277
12	0,256	0,314	0,213	0,275	0,240	0,243
13	0,300	0,108	0,165	0,053	-0,001	0,145
14	0,174	0,295	0,319	0,356	0,386	0,292
15	0,185	0,174	0,208	0,174	0,183	0,184
16	0,234	-0,031	0,428	0,129	0,086	0,200
17	0,272	-0,073	0,224	0,040	0,127	0,136
18	0,215	0,182	-0,002	0,321	0,216	0,193
19	0,136	0,175	0,226	0,129	0,140	0,180
20	0,276	0,177	0,095	0,391	0,326	0,257
21	0,321	0,304	0,127	0,344	0,352	0,282
22	0,560	0,488	0,415	0,315	0,652	0,471
23	0,346	0,527	0,316	0,348	0,622	0,439
24	0,424	0,214	0,241	0,213	0,429	0,303
25	0,288	0,364	0,299	0,179	0,373	0,311
26	0,193	0,170	0,209	0,211	0,379	0,223
27	0,468	0,552	0,336	0,404	0,409	0,453
28	0,064	0,304	0,267	0,181	0,310	0,238
29	0,326	0,139	0,382	0,220	0,175	0,262
30	0,224	0,465	0,419	0,267	0,427	0,382
31	0,339	0,528	0,479	0,411	0,684	0,503
32	0,226	0,422	0,285	0,329	0,588	0,384

33	0,288	0,303	0,348	0,199	0,454	0,334
34	0,128	0,575	0,491	0,407	0,594	0,465
35	0,191	0,593	0,425	0,423	0,639	0,476
36	0,377	0,490	0,285	0,373	0,412	0,385
37	0,394	0,187	0,387	0,252	0,407	0,331
38	0,423	0,313	0,286	0,243	0,387	0,328
39	0,352	0,181	0,079	0,178	0,289	0,224
40	0,445	0,670	0,515	0,562	0,485	0,557
41	0,404	0,481	0,359	0,449	0,624	0,470
42	0,269	0,537	0,374	0,455	0,454	0,435
43	0,457	0,481	0,539	0,529	0,599	0,541
44	0,445	0,349	0,237	0,332	0,053	0,267
45	0,399	0,590	0,412	0,529	0,566	0,513

As correlações item-total mostram que, dos 45 itens que compõem o sistema de avaliação, 10 (22,2%) não discriminaram adequadamente as idades, ou seja, tiveram uma frequência de pontos muito elevada ou muito reduzida, de forma que não contribuem adequadamente para o resultado final do teste. Os itens foram: 2 (Figura 5); 4 (Figura 6); 5, 7 e 8 (Figura 7); 13 (Figura 9); 15 e 17 (Figura 10) e 18 e 19 (Figura 11). Foram realizados novos cálculos do Alfa de Cronbach excluindo estes itens e a alteração no resultado não foi significativa (o alfa passou a ser 0,852) o que parece indicar que sua permanência no teste, não interfere significativamente. Além disso, também foram calculados os alfas caso o item fosse excluído e como pode ser observado na Tabela 52, os resultados também não se modificaram significativamente.

Tabela 52. Variações nos valores de alfa com a exclusão de itens com baixa correlação item-total r_{it}

Itens	Alfa com item excluído
2	0,842
4	0,842
5	0,844
7	0,846
8	0,844
13	0,843
15	0,843
17	0,842
18	0,842
19	0,843

Assim como foi feito para o grupo de 6 e 7 anos, a Tabela 53 mostra a frequência percentual dos pontos dos itens do Koppitz-2 das crianças entre 8 e 12 anos. As comparações foram realizadas por meio do cálculo do Qui-quadrado para verificar se as diferenças discriminavam os grupos em função da idade dos participantes.

Tabela 53. Frequências percentuais de acerto dos itens e qui-quadrado por idade para crianças de 8 a 12 anos

Itens	Idade					Total	Qui-quadrado
	8	9	10	11	12		
1	82,4	85,7	80,7	88,5	84,5	84,0	2,718
2	96,1	85,6	95,8	96,6	98,6	96,4	1,303
3	59,8	67,0	59,7	63,2	60,6	61,9	1,577
4	92,2	95,6	96,6	97,7	95,8	95,5	4,034
5	69,6	63,7	71,4	70,1	64,8	68,3	2,031
6	38,2	40,7	30,3	42,5	38,0	37,4	4,027
7	35,3	42,9	37,8	41,4	49,3	40,6	4,012
8	29,4	45,1	40,3	50,6	45,1	41,5	10,004*
9	31,4	41,8	45,4	40,2	46,5	40,9	5,777
10	31,4	40,7	47,9	37,9	43,7	40,2	6,183
11	71,6	83,5	85,7	83,9	88,7	82,3	11,299*
12	52,9	48,4	50,4	42,5	53,5	49,6	2,721
13	64,7	68,1	72,3	70,1	76,1	70,0	3,045
14	68,6	62,7	73,1	60,9	74,6	68,9	4,819
15	55,9	52,5	57,1	48,3	66,2	57,7	6,326
16	5,9	3,3	12,6	4,6	7,0	7,0	8,603
17	4,9	2,2	6,7	2,3	5,6	4,5	3,746
18	54,9	61,5	60,5	71,3	73,2	63,4	9,022
19	60,8	67,0	70,6	71,3	74,6	68,5	4,698
20	24,5	22,0	20,2	29,9	39,4	26,2	10,281*
21	40,2	47,3	45,4	62,1	42,3	47,2	10,580*
22	38,2	39,6	42,9	47,1	38,0	41,3	2,160
23	23,5	29,7	31,9	40,2	40,8	32,6	8,707
24	72,5	78,0	73,9	83,9	74,6	76,4	4,207
25	77,5	87,9	87,4	92,0	83,1	85,5	9,372
26	61,8	63,7	64,7	69,0	59,2	63,8	1,894
27	30,4	45,1	43,7	49,4	63,4	45,1	19,245**
28	42,2	53,8	57,1	52,9	56,3	52,3	5,888
29	31,4	33,0	40,3	34,5	43,7	36,4	4,130
30	25,5	35,2	31,9	42,5	46,5	35,3	10,762*
31	11,8	20,9	24,4	23,0	28,2	21,3	8,364

32	8,8	9,9	12,6	12,6	22,5	12,8	8,188
33	3,9	3,3	5,9	4,6	11,3	5,5	6,019
34	6,9	17,6	19,3	16,1	19,7	15,7	8,302
35	7,8	15,4	15,1	14,9	22,5	14,7	7,364
36	33,3	36,3	37,8	46,0	40,8	38,5	3,584
37	62,7	56,0	66,4	70,1	81,7	66,6	12,993*
38	61,8	72,5	78,2	66,7	67,6	69,8	7,948
39	89,2	94,5	95,0	93,1	94,4	93,2	3,529
40	21,6	40,7	35,3	46,0	54,9	38,3	23,232**
41	6,9	*17,6	7,6	12,6	18,3	11,9	10,224*
42	7,8	19,8	15,1	20,7	21,1	16,4	8,678
43	8,8	17,6	17,6	31,0	35,9	20,9	24,609**
44	8,8	14,3	3,4	4,6	12,7	8,3	11,490*
45	17,6	31,9	23,5	36,8	39,4	28,7	14,860**

É possível verificar que dos 45 itens, 12 diferenciaram os grupos de forma estatisticamente significativa são eles: 8 (Figura 7); 11 (Figura 9); 20 (Figura 11); 21 (Figura 12); 27 e 30 (Figura 13); 37 (Figura 15) e 40, 41, 43, 44, e 45 (Figura 16). De forma geral a análise das frequências mostra uma tendência a pontuações menores para os grupos de crianças mais novas (8 e 9 anos) quando comparadas com as mais velhas (11 e 12 anos).

Os resultados da análise de itens permitem afirmar que, de forma geral, há uma adequada discriminação dos itens de acordo com a correlação item-total e os resultados do qui-quadrado tanto para o grupo com 6 e 7 anos quanto para os de 8 a 12. A Figura 7 mostra a frequência percentual dos pontos de cada item por idade para 8 a 12 anos.

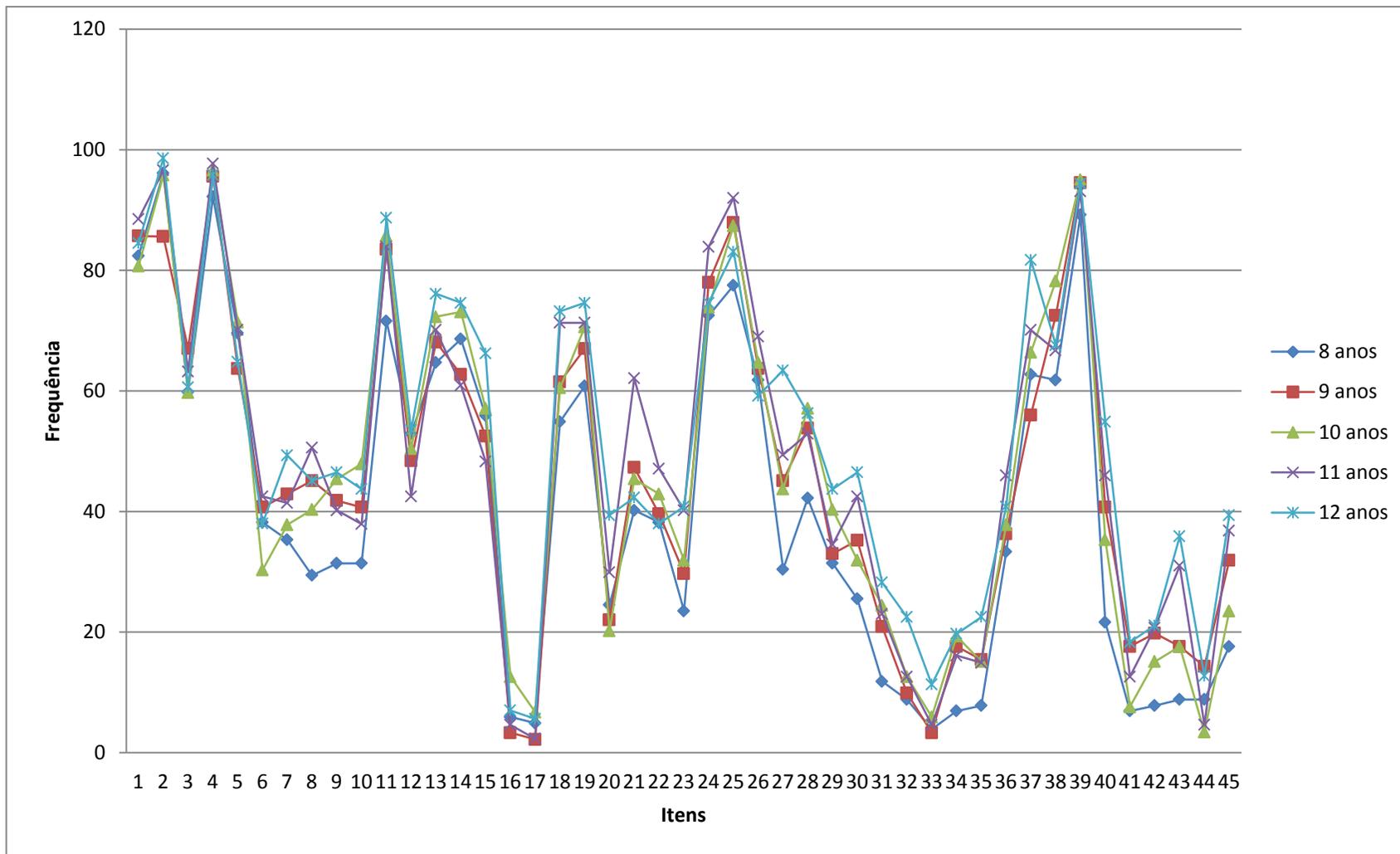


Figura 7. Frequência percentual dos pontos dos itens do Koppitz-2 por idade para o grupo de 8 a 12 anos

V. DISCUSSÃO

O Teste Gestáltico Visomotor de Bender é um instrumento utilizado em avaliações psicológicas para os mais diversos fins tanto em crianças quanto em adultos. Desde sua construção em 1938 realizada por Lauretta Bender foram desenvolvidos diferentes sistemas de avaliação (Koppitz, 1963/1989; Santucci-Granjon, 1968; Clawson, 1962/1992; Hutt, 1969; Brannigan & Decker, 2003; Reynolds, 2007 e outras).

Brannigan e Decker (2003) realizaram uma revisão do Teste de Bender, no qual adicionaram novos cartões, quatro mais fáceis e três mais difíceis para aumentar a amplitude de idades para utilização do teste, totalizando 16 modelos. Suas pesquisas abrangeram uma amostra de 3.535 participantes com idades entre 4 e 85 anos. No entanto, a proposta destes autores foi criar um sistema de avaliação mais qualitativo, diferente do sistema de avaliação de Koppitz (1963/1989), que é um dos mais usados para este teste. O Bender Koppitz-2 (Reynolds, 2007) foi proposto como uma revisão do sistema de avaliação original de Koppitz, ampliado com os novos modelos elaborados por Brannigan e Decker (2003). Para o Koppitz-2 foram desenvolvidos dois critérios diferentes de avaliação, um para crianças de 5 a 7 anos, constituído por 34 itens e outro, para a faixa de 8 a 85 anos, com 45 itens. Os 16 modelos da nova versão do teste também são divididos em função da idade, ou seja, nos mais novos são aplicados os modelos de 1 a 13 e nos mais velhos os de número 5 a 16. Nesta nova proposta, todos os examinandos realizam as nove figuras originais do teste, sendo que foram incluídas três mais fáceis antes dos nove originais para a faixa de 5 a 7 anos e quatro mais difíceis que são aplicados depois dos cartões originais na faixa de 8 a 85 anos. Além das diferenças no número e no tipo de cartões administrados aos dois grupos em função da idade, também há diferenças nos critérios de avaliação em alguns itens dos desenhos que são comuns às duas faixas etárias.

Reynolds e Fletcher-Janzen (2000) definem a integração visomotora como a capacidade para relacionar o estímulo visual e a resposta motora de forma precisa e adequada e está mais relacionada a tarefas que envolvem a interpretação ou organização de elementos físicos do estímulo do que a aspectos simbólicos do mesmo. Segundo Reynolds (2007) alguns pesquisadores consideram o Bender um teste de percepção visual e outros como de coordenação motora, mas para Koppitz

(1975, citado por Reynolds 2007) ele não seria nenhum dos dois, mas sim um teste de integração visomotora, que envolve níveis de processos de integração superiores. Afirma que, ainda que exista uma dificuldade para separar as capacidades perceptual e motora na reprodução das figuras do Bender, a função gestáltica inerente na integração visomotora é significativamente mais complexa do que a percepção ou a coordenação motora tomadas isoladamente.

Para Reynolds (2007) o Koppitz-2 tem cinco usos principais: 1) constatar a presença e o grau de dificuldades visomotoras nos examinandos; 2) identificar candidatos a encaminhamentos para outros tratamentos e diagnósticos mais específicos; 3) verificar a efetividade de programas de intervenção desenvolvidos para corrigir problemas visomotores, que podem interferir não somente nos processos de aprendizagem, mas na vida de pessoa como um todo; 4) como instrumento de pesquisas e 5) colaborar em diagnósticos diferenciais de diversas condições psicológicas e neuropsicológicas ao longo da vida. A ênfase do sistema de pontuação do Koppitz-2 é avaliar a progressão do desenvolvimento da capacidade de eliminar erros específicos na reprodução de figuras.

A avaliação do Koppitz-2, segundo Reynolds (2007) consiste primeiramente em obter a pontuação bruta por meio da contagem do total de pontos de acordo com os critérios de avaliação para cada faixa etária. O total de pontos é então convertido para percentil e/ou para um Escore Padrão ($M=100$; $DP=15$), que é chamado de Índice Viso-Motor (Visual Motor Index, VMI – em inglês). As tabelas de conversão foram organizadas em função da idade: dos 5 aos 16 anos as tabelas foram elaboradas com intervalos de três em três meses; dos 17 aos 20 são anuais e acima dos 20 anos são divididas em faixas de 5 anos até a idade de 85 anos ou mais velhos. As faixas de classificação foram elaboradas de acordo com os resultados do VMI (Menor do que 70: Significativamente prejudicado; entre 70-79: Mediana a Moderadamente prejudicado; 80-89: Abaixo da Média; 90-109: Média; 110-119: Acima da Média; 120-129: Superior e Acima de 130: Muito Superior). Reynolds ressalta a importância do uso do VMI em conjunto com o percentil para a interpretação do teste. O manual oferece ainda normas em Notas T ($M=50$; $DP=10$), Z-Escore ($M=0$; $DP=1$), Equivalente da Curva Normal - NCE ($M=50$; $DP=21,06$) e Estanino ($M=5$; $DP=2$). Na presente pesquisa, optou-se por calcular as normas somente em percentil em função da familiaridade dos psicólogos brasileiros com este tipo de norma. O autor ressalta ainda a importância de se considerar o tempo

de aplicação na avaliação do teste, apesar de não apresentar tabelas específicas para avaliá-lo quantitativamente.

Outro parâmetro de avaliação fornecido no manual são os Equivalentes de Idade (*Age Equivalents*), que são elaborados a partir da média de pontos brutos obtida no teste pelas crianças para cada faixa de idade. Reynolds (2007) apresenta equivalentes de idade para as faixas de 5 a 18 anos apenas, porque considera que em variáveis relacionadas ao desenvolvimento, este tipo de parâmetro apresenta pouco significado psicológico ou psicométrico em examinandos com idades maiores. No entanto, aponta que o uso deste tipo de medida tem sido amplamente questionado na literatura científica por apresentarem propriedades estatísticas inadequadas e estarem repletos de problemas de interpretação. Assim, apesar de apresentados no manual, não são recomendados, devendo ser privilegiadas as interpretações dos escores padronizados. Por este motivo, optou-se também por não calcular este parâmetro na pesquisa brasileira com o teste.

No levantamento bibliográfico realizado foram encontradas apenas três referências com o Koppitz-2, uma delas faz somente uma apresentação do teste não sendo, portanto um relato de pesquisa. As outras duas são de um resumo do *Dissertation Abstracts International* e um artigo científico, ambos sobre pesquisas mais específicas voltadas para a área clínica, todas fruto de publicações internacionais. Dessa forma há poucos artigos com a descrição de pesquisas com o Koppitz-2, mesmo fora do Brasil, por este motivo, a discussão será conduzida em relação às pesquisas que constam no manual do teste (Reynolds, 2007) e dos estudos do mesmo tipo dos apresentados neste trabalho realizados com os cartões originais e com outros sistemas de avaliação.

Os primeiros resultados apresentados neste trabalho tiveram a finalidade de mostrar como foi o desempenho no Koppitz-2 de crianças de diferentes escolas públicas da cidade de São Paulo. A amostra se restringiu a este grupo em função de dificuldades de acesso às escolas particulares para a realização da pesquisa. Foram feitas Análises de Variância para comparar as médias dos grupos em função da idade, do sexo e do tipo de escola, bem como as interações entre as variáveis. As diferenças entre as médias de pontos foram estatisticamente significantes em relação à idade das crianças do grupo de 6 e 7 anos, com média de 13,76 (DP=5,25) para 6 anos e de 17,31 (DP=4,24), para 7 anos. Este resultado mostra que ocorreu um efeito do desenvolvimento, ou seja, as crianças mais velhas tenderam a realizar

os desenhos com melhor qualidade. Em relação ao grupo de 8 a 12 anos também ocorreram diferenças entre as idades, embora o *Teste de Post Hoc de Tukey* tenha mostrado que a diferença significativa somente foi encontrada entre a média de pontos de 8 anos, quando comparada com às demais idades. Neste grupo também foi identificado um aumento da pontuação com a progressão da idade, o que reforça a hipótese do efeito do desenvolvimento na melhor reprodução dos desenhos.

Reynolds (2007) fez uma extensa descrição da amostra de padronização americana do Koppitz-2, apresentando a distribuição de frequência em função da idade, sexo, raça, região geográfica e nível sócio-econômico. Ele informou ainda, que usou a mesma amostra da pesquisa de Brannigan e Decker (2003), coletada para obtenção das normas do Teste Gestáltico Viso-motor de Bender segunda edição ou Bender 2. Apresenta também a descrição dos grupos clínicos que participaram da pesquisa (deficientes mentais; pessoas com dificuldades específicas de aprendizagem: que estão relacionadas a distúrbio em um ou mais dos processos psicológicos básicos envolvidos na compreensão da linguagem, que resulta numa incapacidade de ouvir, pensar, falar, ler, escrever, soletrar ou fazer cálculos matemáticos; TDAH; distúrbios emocionais graves; autismo e superdotados).

No entanto, apesar da extensa descrição da amostra que é feita no manual, não há informações sobre as médias das pontuações dos diferentes grupos, o que dificultou a realização de comparações entre os estudos originais e os do presente trabalho para verificar, se as eventuais diferenças entre as médias da padronização americana e as desta pesquisa são ou não estatisticamente significantes. A Figura 5 apresentada nos resultados ilustra uma tentativa de comparação entre as médias das diferentes amostras de normatização. Uma vez que o manual apresenta somente uma tabela de avaliação geral em função da idade, a Figura 5 também foi construída considerando a amostra total brasileira. Optou-se por usar os pontos correspondentes ao percentil 50, porque este pode ser considerado o mais próximo das médias dos grupos. As pontuações da amostra brasileira foram: 14, 18, 17, 20, 21, 22 e 22 e as americanas: 13, 15, 16, 18, 20, 22 e 24, respectivamente para as idades 6 a 12 anos.

Rueda, Bartholomeu e Sisto (2006) compararam os resultados das pontuações do Bender Koppitz em função da idade. O Bender foi avaliado pelos indicadores Maturacionais (agrupados em quatro categorias: distorção da forma, integração, rotação e perseveração). A amostra foi composta por 312 crianças,

51,6% do sexo masculino e 48,4% do feminino, com idades entre 7 e 10 anos (média de 8,62 anos e desvio padrão de 1,12), alunos do primeiro ao quinto ano do ensino fundamental de uma escola pública do interior de São Paulo. No estudo, os autores somente consideraram os itens referentes à distorção de forma e à integração, e um escore total obtido pela soma dos pontos dos itens referentes a estas categorias, que poderia variar entre 0 e 11. As médias foram comparadas por meio de análise de variância e os resultados mostraram diferenças estatisticamente significantes entre as idades ($F=11,76$; $p<0,01$), confirmando as colocações de Koppitz (1963/1989) de que à medida que ocorre o desenvolvimento cognitivo da criança, há uma tendência à diminuição dos pontos no Bender (lembrando que neste sistema quanto maior a pontuação pior é o resultado do examinando, indicando desenhos com maior número de distorções em relação aos modelos, diferentemente do Koppitz-2 em que são pontuados os acertos).

Na presente pesquisa para a faixa de 6 e 7 anos também foi encontrada diferença significativa entre as médias em função do sexo, com a média das meninas (16,57 pontos, DP=4,88) maior do que a dos meninos (14,91 pontos, DP=5,05). Para a faixa de 8 a 12 anos, a diferença significativa ocorreu em relação ao tipo de escola (com a média maior para as escolas municipais) e para a interação entre a idade e o sexo. Os resultados indicaram a necessidade de elaboração de tabelas separadas por sexo e idade para as crianças de 6 e 7 anos e, para os de 8 a 12 anos, por idade, tipo de escola e por idade e sexo.

Em relação ao tempo de aplicação os resultados foram maiores para as crianças com 6 anos, quando comparadas às de 7, respectivamente 10,57 minutos (DP=5,82) e 8,50 minutos (DP=3,27). Em relação ao grupo de 8 a 12 anos somente o tipo de escola diferenciou significativamente as médias, com média menor para a escola estadual.

A validade é uma das mais importantes propriedades psicométricas de um teste, pois é ela que proporciona segurança nas interpretações que serão realizadas a partir dos seus resultados (Anastasi e Urbina, 2000; Urbina, 2007; Cohen et al., 2014). De acordo com o *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA] & National Council on Measurement in Education [NCME], 1999), a validade de um teste indica o grau em que as evidências e a teoria sustentam as interpretações pretendidas a partir das respostas dadas pelos examinandos ao teste.

A AERA, APA e NCME (1999) distingue cinco tipos de evidências de validade: 1) Baseada no conteúdo do teste; 2) Baseada nos processos de resposta; 3) Baseada na estrutura interna; 4) Baseada na relação com outras variáveis e 5) Baseada nas consequências da testagem. A partir desta categorização, pode-se afirmar que os estudos realizados no presente trabalho se encaixam no tipo 4, que se refere à existência de relações entre os pontos no teste e variáveis externas, tais como as comparações de grupos contrastantes (clínico e controle), variáveis relacionadas ao desenvolvimento e correlações entre os escores do teste e outras variáveis que avaliem o mesmo construto ou construtos relacionados (convergência) e com variáveis medindo construtos diferentes (divergência). Podem ser incluídos, ainda, pesquisas sobre a capacidade preditiva do teste e outros critérios externos, que de alguma forma se associam à finalidade direta do teste.

Foram realizados os seguintes estudos de validade do Koppitz-2: 1) relação entre instrumentos que avaliam o mesmo construto, ou seja, a percepção visual e a capacidade de integração visomotora (validade convergente) no caso com o teste das Figuras Complexas de Rey; 2) relação entre testes que avaliam construtos relacionados por meio da relação com o teste de Inteligência Não Verbal - R-2 e entre os sistemas de avaliação de Koppitz e Koppitz-2 e 3) Validade relacionada ao desenvolvimento, pela comparação entre os desempenhos no teste de crianças de diferentes faixas etárias (este último já foi discutido acima ao abordar as comparações do estudo normativo).

O Teste das Figuras Complexas de Rey pode ser usado em avaliações de algumas funções neuropsicológicas, tais como a percepção visual (como a pessoa apreende as informações e dados perceptivos), a memória imediata e a integração visomotora (Oliveira & Rigoni, 2010). Uma vez que a tarefa do examinando consiste em copiar as figuras apresentadas (Figura A ou B em função de sua idade), é preciso ser capaz de perceber os detalhes da figura para conseguir copiá-la adequadamente. Ainda que os desenhos tanto da Figura A quanto da Figura B sejam muito mais complexos que os do Bender, é possível considerar, que as tarefas são bastante semelhantes, caracterizando que a avaliação da relação entre o desempenho nos dois testes indique a validade para o Koppitz-2. Os resultados mostraram correlações positivas e significantes de magnitude moderada para as crianças de 6 e 7 anos (0,639) e elevada para os de 10 e 11 (0,720), de acordo com os parâmetros apresentados por Dancey e Reidy (2013), indicando uma elevada

relação entre os dois instrumentos se caracterizando, portanto como evidência de validade para o teste (Guilford, 1950; Nick & Kellner, 1971; Fonseca, 2003 e Nunes & Primi, 2010).

Outra questão a ser observada é o aumento da correlação com o crescimento da idade, indicando uma tendência para realizar os desenhos tanto do Bender quanto das Figuras Complexas com maior precisão à medida que progride a idade.

Reynolds (2007) cita um estudo realizado com outro instrumento de avaliação da integração visomotora o *Beery-Buktenica Developmental Test of Visual-Motor Integration – Fourth Edition*. A amostra foi formada por 45 examinandos entre 5 e 16 anos, sendo 57,8% do sexo masculino e 42,3% do feminino. Os resultados não foram descritos em função das idades e as correlações, corrigidas para restrição de faixa e por eventual falta de precisão do *Beery-Buktenica*, foram 0,46 e 0,51, respectivamente, um pouco mais baixas do que as desta pesquisa.

Por definição, o fator geral de inteligência (fator *g*) apresentado por Spearman (1955) é um fator comum presente em todas as capacidades humanas. O fator *g* recebeu este nome “*porque, ainda que varie livremente de um indivíduo para os outros, permanece inalterável em um mesmo indivíduo, em relação a todas as demais capacidades correlacionadas*” (Spearman, 1955, p.82). Rosa e Alves (2012a) afirmam que o fator *g* é um fator quantitativo e está presente em todas as capacidades cognitivas do homem e por este motivo pode-se levantar a hipótese da possibilidade da existência de uma relação entre o Koppitz-2 e o R-2, justificando este objetivo desta pesquisa, o que também é sustentado pelos diversos estudos encontrados na literatura que relacionam a integração visomotora e as capacidades cognitivas (Koppitz, 1963/1989; Rueda et al., 2006; Reynolds, 2007; Bartholomeu & Sisto, 2008; Sisto, Bartholomeu, Rueda, Santos & Noronha, 2008; Rosa & Alves, 2012b, entre outros).

Segundo Reynolds (2007), a partir da reprodução dos desenhos o Koppitz-2 avalia a habilidade de integração visomotora global que pode ser definida como a capacidade de relacionar os estímulos visuais com respostas motoras. Para este autor, estas características avaliadas no teste apresentam relação e podem ser afetadas pelas capacidades viso-perceptivas, motoras finas e outras habilidades, incluindo àquelas que se relacionam com a inteligência. No entanto, para a maior parte das crianças e adultos a integração visomotora é a característica mais evidente do Koppitz-2. Dessa forma, segundo ele, o Koppitz-2 deverá apresentar maiores

relações com medidas de inteligência não verbal que envolvam tarefas que exigem capacidades visomotoras de diversos tipos do que aquelas relacionadas às fluídas em que a atividade motora esteja menos presente. Sendo assim o autor concluiu que o Koppitz-2 deve apresentar uma associação maior com as medidas de inteligência fluída não verbal do que com as medidas verbais ou cristalizadas da inteligência.

Koppitz (1963/1989) estudou a relação entre a inteligência e a maturidade percepto-motora em 239 crianças (176 normais e 63 deficientes) de 5 a 10 anos, que eram atendidas em uma clínica de orientação infantil para avaliação por causa de problemas emocionais, dificuldades de aprendizagem ou ambas. Aplicou o Bender e o Stanford-Binet Forma L (1937) para crianças menores de 7 anos ou o WISC (1949) para as de 7 anos ou mais e para as deficientes com idade mental de no mínimo 7 anos, para determinar o nível intelectual. As crianças foram organizadas em dois grupos em função dos resultados dos testes de Inteligência (um grupo com QI médio de 98 e outro de 63) e as pontuações obtidas nos dois testes foram correlacionadas. Encontrou correlações negativas entre -0,48 e -0,79 ($p \leq 0,01$), sendo a maior para 5 anos. A correlação entre o Bender e a idade mental para os deficientes foi de -0,85. A autora sugere que o teste poder ter um papel importante como um instrumento de seleção de crianças que estão ingressando na escola. Verificou ainda que o desempenho no Bender está relacionado a cinco dos subtestes do WISC: Aritmética, Completar Figuras, Arranjo de Figuras, Cubos e Armar Objetos. Koppitz concluiu que o Bender pode ser utilizado como uma estimativa da inteligência das crianças entre 5 e 10 anos.

Kroeff (1988,1991) também realizou uma pesquisa com a finalidade de estabelecer normas de avaliação do Koppitz em crianças de escolas públicas e particulares. O autor encontrou diferenças significativas entre as médias em função das séries e das idades ainda que não tenha informado quais as idades foram de fato discriminadas.

No estudo de Rueda et al. (2006), mencionado anteriormente, os autores também investigaram a relação entre a maturidade perceptual e a inteligência. Além do Bender (avaliado pelo sistema de Koppitz original) as crianças também fizeram o Teste da Figura Humana como medida de inteligência, sendo avaliado pelos critérios de Goodenough. As correlações entre as pontuações dos dois testes foram baixas e negativas, mas estatisticamente significantes (variaram entre -0,16 e -0,21). Os

autores concluíram que estes resultados confirmam a tendência de que, quanto melhor o desempenho no Desenho da Figura Humana (no caso a medida da inteligência), menor será a pontuação no Bender (pontuado pelos erros). É importante ressaltar que o fato de não terem sido usados todos os itens de Koppitz, mas somente os itens de distorção da forma e integração, pode ter influenciado os resultados colaborando assim para que as correlações fossem mais baixas.

Em outro estudo, Bartholomeu e Sisto (2008) também investigaram a relação entre a inteligência e a maturidade visomotora. A amostra foi composta por 244 alunos metade de cada sexo, com idades entre 7 e 10 anos (média de 8 anos e $DP=1,08$), cursando do primeiro ao quinto ano do ensino fundamental em uma escola pública do interior de São Paulo. Foram aplicados coletivamente o Desenho da Figura Humana – Escala Sisto (Sisto, 2005) e o Teste de Bender Sistema de Pontuação Gradual (Sisto et al., 2005). Os testes foram avaliados de acordo com as instruções dos respectivos manuais e calculadas correlações entre os dois testes por sexo, idade, para a interação entre estas variáveis e para a amostra total. As correlações foram negativas, de magnitude moderada e estatisticamente significantes ($p<0,01$), tanto para a amostra total (-0,56), quanto por sexo (-0,54 para o masculino e -0,58 para o feminino), indicando uma tendência de que, quanto maior a pontuação no DFH, menor a do Bender, ou seja, quanto melhor e mais detalhado o Desenho da Figura Humana, menor o total de erros no Bender. As correlações relativas às idades também foram significantes (variaram entre -0,25 e -0,55, para $p \leq 0,05$). As correlações foram mais altas aos 7 anos (-0,55), seguida pela de 9 anos (-0,46), pela de 8 (-0,36) e por último a menor de 10 anos (-0,25). Os autores concluíram que os dados mostram uma tendência de que quanto maior a inteligência, maior a maturidade visomotora, o que também foi observado nas diferentes idades, ainda que em graus diferentes, portanto confirmando a relação entre os construtos.

Sisto et al. (2008) pesquisaram a relação entre o Teste de Bender avaliado pelo Sistema de Pontuação Gradual (B-SPG) e o Teste de Matrizes Coloridas de Raven. A amostra foi formada por 239 crianças entre 7 e 10 anos (média de 8 anos e $DP=1,11$), que foram avaliadas com os dois testes coletivamente. A escolaridade variou entre o primeiro e o quinto anos do ensino fundamental, 55,2% eram meninas e 44,8% meninos. As correlações entre o total de pontos do Raven e por série do teste, com e sem o controle das idades, variaram entre -0,46 e -0,55 sem o controle

da idade e entre -0,31 e -0,38 com este controle. Os autores afirmaram que as crianças mais inteligentes tendem a ter um menor número de distorções ao copiarem os desenhos do Bender. Investigaram também a interferência da série escolar e dos resultados extremos no Raven e verificaram que em geral os resultados confirmaram esta tendência, ainda que algumas correlações entre as séries indicassem flutuações que podem sugerir que, em alguns momentos do desenvolvimento, as relações entre a maturidade percepto-motora e a inteligência podem estar mais relacionadas do que em outros.

Flores-Mendoza, Mansur-Alves, Abad e Lelé (2010) também estudaram a relação entre os resultados do Bender com o R-2 e o Desenho da Figura Humana em 293 crianças de escolas públicas e particulares de Belo Horizonte, (157 meninos e 136 meninas), com idade média de 8,13 anos (DP=1,88) e amplitude entre 5 e 11 anos. As correlações entre o DFH e o Bender variaram entre -0,391 e -0,508 e entre o R-2 e o Bender entre -0,287 e -0,332, todas significantes ($p<0,05$), indicando uma relação baixa a moderada entre as variáveis (Dancey & Reidy, 2013).

Rosa e Alves (2012b) pesquisaram a relação entre a inteligência e a organização visomotora infantil utilizando o Teste de Bender (avaliação Koppitz) e o Teste de Inteligência Não Verbal R-2. Participaram deste estudo 538 crianças com idades entre 4 anos e 9 meses e 11 anos e 9 meses que foram avaliadas pelos dois instrumentos. A correlação entre os totais de pontos dos dois testes foi de -0,702 ($p<0,01$) considerada forte (Dancey & Reidy, 2013). As autoras apresentaram também os resultados por idade (variaram entre -0,139 e -0,699), por sexo dos participantes (feminino -0,722 e masculino -0,682) e por tipo de escola (-0,701 municipais e -0,656 particulares). Somente não foi significativa a correlação para 5 anos (0,139), para as demais idades foi superior a -0,426 e a maior foi para 11 anos (-0,699), estes resultados revelam uma relação significativa entre as capacidades medidas pelo R-2 e o Bender.

Reynolds (2007) relata alguns estudos de validade realizados com o Koppitz-2 com objetivo relacionar os seus resultados com testes de inteligência. Foram obtidas correlações com os testes WISC-III, WAIS-III e Stanford-Binet V. A amostra do WISC-III foi constituída por 58 participantes com idades entre 6 e 16 anos. As correlações (corrigidas para restrição de idades e atenuação da precisão) foram maiores com o QI de Execução (0,63) em relação ao QI Verbal (0,34) e ao Total (0,56). Em relação aos índices fatoriais, as correlações mais elevadas foram

para a Organização Perceptual (0,64) e Resistência à Distração (0,46), seguidas por Compreensão Verbal (0,29) e Velocidade de Processamento (0,26). Dessa forma, os resultados apontam correlações mais elevadas nos subtestes do WISC-III que exigem capacidades visomotoras e não verbais, confirmando os dados de Koppitz (1963/1989).

O estudo com o WAIS-III foi conduzido com uma amostra de 80 adultos com idades entre 16 e 84 anos. Como apontado por Koppitz (1963/1989), a relação entre o desenvolvimento da integração percepto-motora e a inteligência diminui à medida que a idade aumenta, por isso as correlações com o Bender nos adultos foram menores e ficaram entre 0,36 e 0,38 para os QIs e 0,29 e 0,36 para os Índices Fatoriais. As mais elevadas também ocorreram com o QI de Execução (0,38) e com o Índice de Organização Perceptual (0,36) confirmando o que já foi mencionado em relação ao WISC-III, ou seja, uma relação maior nas provas que exigem capacidades visomotoras.

A comparação com o Stanford-Binet V foi feita com 387 participantes, com idades entre 5 e 84 anos, Para os QIs as correlações variaram entre 0,22 e 0,40, sendo a mais baixa para o QI Verbal e o mais alto para o Não Verbal. Para os índices fatoriais variaram entre 0,26 e 0,39. Reynolds (2007) concluiu que os resultados evidenciam uma relação não muito alta entre a integração visomotora e a inteligência e que ela é mais elevada em tarefas não verbais e que evoluam a capacidade visomotora.

Nesta pesquisa com o Teste de Inteligência Não Verbal – R2, as correlações foram moderadas tanto para 6 e 7 anos (0,594) quanto para 10 e 11 (0,539). As correlações dos meninos foram maiores do que as das meninas, ainda que a diferença não tenha sido muito grande. Quando separadas pela idade a correlação maior foi aos 11 anos (0,667), seguida por 6 anos (0,638), depois 7 anos (0,523) e a mais baixa para os 10 anos (0,439).

Para Reynolds (2007), a associação entre o Bender e a Inteligência é maior em tarefas não verbais, o que foi confirmado nos estudos por ele descritos e também pelo presente estudo, que apesar de não utilizar um teste verbal para comparação, apresentou resultados próximos aos apontados por ele no QI e nos índices fatoriais que exigem a integração visomotora (a correlação com o QI de Execução foi de 0,63 e o Índice de Organização Perceptual 0,64 enquanto as correlações com o R-2 foram 0,594 e 0,539).

Na comparação dos resultados desta pesquisa com os demais estudos citados deve-se observar que as correlações com os outros sistemas foram todas negativas, uma vez que neles são pontuados os erros e para o Koppitz-2 são os acertos. As correlações das duas faixas etárias desta pesquisa foram semelhantes aos apresentados por Bartholomeu e Sisto (2008), Sisto et al. (2008), maiores do que os de Flores-Mendoza et al. (2010) e um pouco menores do que os de Rosa e Alves (2012b). A Tabela 60 mostra a comparação das correlações nos diferentes estudos.

A Tabela 60. Correlações por idade nas diferentes pesquisas sobre a relação entre o Bender e a inteligência

Idade	Atual	Koppitz (1963/1989)	Bartholomeu e Sisto (2008)	Flores-Mendoza et al. (2010) DFH e Bender / R- 2 e Bender	Rosa & Alves (2012b)
6	0,638	-0,63	-	-0,508/-0,319	-0,431
7	0,523	-0,61	-0,55		-0,497
10	0,439	-0,59	-0,25	-0,425/-0,316	-0,479
11	0,667	-	-		-0,699

Pode-se verificar que para 6 anos, os resultados da pesquisa atual e a de Koppitz são iguais, enquanto que para as de Flores-Mendoza et al. (2010) e Rosa e Alves (2012b), ambas com o R-2, foram mais baixos. Um dos fatores que pode explicar o resultado bem mais baixo no estudo de Flores-Mendoza et al. (2010) é que os autores juntaram no mesmo grupo as crianças de 5 a 7 anos, o que eventualmente pode ter interferido nos resultados. Em relação a pesquisa de Rosa e Alves (2012b) a amostra foi proveniente de uma cidade do interior e teve participantes de escolas públicas e particulares, o que não ocorreu na pesquisa atual. Os resultados para as crianças de 7 anos foram muito próximos em todos os estudos descritos. Para 10 anos, as correlações foram menores em todas as pesquisas sendo que no estudo de Bartholomeu e Sisto (2006) elas foram mais baixas do que as demais. Diferentemente de Koppitz (1963/1989) que obteve correlações mais altas para as crianças menores tanto nesta pesquisa como na de Rosa e Alves (2012b) as correlações maiores foram das crianças mais velhas (deve-se considerar, no entanto que a correlação de 11 anos desta pesquisa pode ter sido influenciada pelo número reduzido de participantes (N=16). No entanto, quanto as faixas de 6 e 7 e de 10 e 11 anos, verifica-se que as correlações foram mais

elevadas com os mais novos, mantendo-se a premissa de que com o aumento da idade e com a maturação percepto-motora a associação com a inteligência tende a diminuir.

A Tabela 61 mostra a comparação das correlações por sexo em diferentes pesquisas, nas quais os resultados foram muito próximos em todos os estudos descritos, sendo que na pesquisa de Rosa e Alves (2012b) eles foram um pouco maiores.

A Tabela 61. Correlações por sexo nas diferentes pesquisas

Sexo	Atual		Rosa & Alves	Bartholomeu e Sisto
	6 e 7	10 e 11	(2012)	(2008)
Feminino	0,542	0,520	-0,722	-0,58
Masculino	0,620	0,568	-0,682	-0,54

Entre as hipóteses que podem explicar as diferenças observadas nas correlações pode-se citar o fato de que foram utilizados sistemas de avaliação diferentes tanto para o Bender quanto para o DFH nos diferentes estudos. As correlações entre os sistemas Koppitz e Koppitz-2 (apresentadas nos resultados) apontam nesta direção uma vez que elas foram moderadas, mas não elevadas, o que pode sugerir que as diferenças nas correlações possam realmente ter sido influenciadas pela diferença nos sistemas. Cabe ressaltar, ainda, o fato de que foram diferentes avaliadores que fizeram as avaliações o que também pode ter causado um efeito nos resultados, assim como as variáveis regionais e socioeconômicas dos examinandos.

Koppitz (1963/1989) ressalta que a relação entre a inteligência e a maturidade visomotora pode ser percebida especificamente nas crianças. Uma vez desenvolvida esta maturidade o Bender não serviria mais como um instrumento para avaliar a habilidade intelectual. Segundo Bender (1938, citado por Koppitz, 1963/1989) a cópia das figuras está relacionada com o nível de maturidade visomotora e a percepção visomotora guarda uma estreita relação com a linguagem e outras funções associadas à inteligência em crianças pequenas o que não acontece com os adultos em que esta maturidade já teria atingido seu estágio mais alto de desenvolvimento. Dessa forma quando a pessoa atinge uma idade em que a percepção visomotora amadurece por completo, o Bender não pode ser mais

utilizado como medida de aptidão intelectual. Neste contexto, somente há sentido em realizar pesquisas com o objetivo de relacionar o Bender e a inteligência em crianças. Os resultados apresentados na presente pesquisa confirmam, portanto a evidência de validade entre o Koppitz-2 e o R-2.

A correlação entre a avaliação de Koppitz e o Koppitz-2 foi calculada considerando todas as figuras de cada faixa etária e somente os comuns às duas avaliações. Em todos os casos (com exceção da idade de 9 anos), as correlações foram maiores, quando são considerados os modelos comuns aos dois sistemas. Quando se consideram os resultados por faixa de idade, em geral, os resultados dos mais novos foram maiores do que os mais velhos e dos meninos maiores do que os das meninas. Para a amostra dos mais novos a correlação com todas as figuras aplicadas foi de -0,647 e para as comuns -0,646, já para os mais velhos as correlações foram -0,557 e -0,566 respectivamente para as figuras aplicadas para esta idade e para as nove figuras comuns. Reynolds (2007) comparou os dois sistemas com uma amostra de 50 crianças com idades entre 5 e 11 anos provenientes do arquivo de uma clínica de avaliação neuropsicológica. Os protocolos foram pontuados nos dois sistemas (somente os modelos comuns) e as correlações encontradas foram de -0,88 sem a correção da idade e -0,73 com esta correção. Portanto, pode-se verificar que as correlações foram um pouco maiores do que as da pesquisa atual. A porcentagem da variância compartilhada no estudo de Reynolds foi de 53,3%, enquanto na pesquisa brasileira foi de 41,7% para os mais novos e 32% para os mais velhos, o que indica que apesar de existir uma relação satisfatória entre os dois sistemas, eles não são semelhantes. Dessa forma, eventuais diferenças entre pesquisas que utilizarem um ou outro sistema podem também ser influenciadas por este fator. As diferenças entre as duas pesquisas também podem ter sido causadas pelo fato de terem sido pontuadas por diferentes avaliadores.

A fidedignidade do Koppitz-2 foi obtida por três métodos: teste e reteste, consistência interna e precisão entre avaliadores. Pelo teste-reteste foi obtido um bom nível de estabilidade temporal para o Koppitz-2: para 6 e 7 anos, a correlação foi de 0,794 e para os de 10 e 11 anos, de 0,850, que podem ser considerados como coeficientes de precisão adequados (Guilford, 1950 e Anastasi & Urbina, 2000). Eles foram mais elevados para as idades de 6 e 11 anos (respectivamente 0,837 e 0,964), do que os de 7 (0,750) e 10 anos (0,745), ainda que os coeficientes de 11

anos devam ser considerados com cautela em função do tamanho do grupo (N=12). Para os dois grupos foram observadas diferenças estatisticamente significantes entre as médias dos totais de pontos no teste e no reteste sendo que as médias foram maiores na segunda aplicação (6 e 7 anos: média do teste de 15,45 pontos e DP=4,11 e do reteste 16,57 pontos e DP=3,91; $t=-3,091$; $p<0,01$; 10 e 11 anos: média do teste de 20,68 pontos e DP=6,60 e do reteste 21,84 pontos e DP=6,39; $t=-2,30$; $p<0,05$). Em relação ao tempo de aplicação, as diferenças também foram estatisticamente significantes, sendo que o intervalo de tempo para completar as figuras no reteste foi menor (para 6 e 7 anos o tempo médio foi de 8,20 minutos e para o reteste 7,08; para os mais velhos, 11,09 no teste e 9,89 no reteste). As médias de pontos e do tempo utilizado para a realização do teste e do reteste parecem confirmar a existência de um efeito de aprendizagem, considerando que o intervalo médio de tempo entre as aplicações (35,71 dias) parece ser reduzido para indicar um efeito do desenvolvimento.

Reynolds (2007) cita uma pesquisa de teste e reteste realizada com o Koppitz-2, com uma amostra de 202 participantes de ambos os sexos, com idades entre 5 e 85 anos que fizeram o teste duas vezes, com um intervalo de médio de 18 dias entre as aplicações (variando entre 14 e 21 dias). Os resultados foram agrupados em função da amplitude das faixas etárias. As correlações foram muito próximas as da presente pesquisa, ainda que as faixas de idades consideradas nos dois estudos sejam diferentes e não esteja claro se foi usado o escore bruto ou os padronizados (VMI) para os cálculos. Para as idades entre 5 e 7 anos a correlação foi de 0,75 e para 8 a 17 anos, 0,82.

Em relação à comparação das médias, Reynolds (2007) relatou que, diferentemente do que encontrado no presente estudo, não foram observadas diferenças estatisticamente significantes. No entanto, esta informação foi apenas referida, pois não foram apresentados no manual os resultados dos testes estatísticos de comparação de médias. As médias de 5 a 7 foram 102 (DP=16) para o teste e 96 (DP=13) para o reteste (escores em VMI) e para os participantes de 8 a 17 anos, 104 (DP=13) e 102 (DP=14) respectivamente para o teste e o reteste. O autor afirma que as diferenças entre as médias foram pequenas, o que sugere que as variações de escores observados no Koppitz-2, com intervalos de até três semanas, podem ter sido causadas por fatores diferentes da prática obtida na primeira aplicação. No entanto, as médias para a presente pesquisa também foram

muito próximas (aproximadamente um ponto em média para cada um das faixas), mas, foram significantes o que pode sugerir que os resultados de Reynolds (2007) podem ter sido influenciados pelo tamanho da amostra em cada faixa de idade, o que não foi informado no manual. Além disso, deve-se considerar as diferenças entre as faixas etárias das duas pesquisas: enquanto neste trabalho foram utilizados os resultados de crianças de 6 e 7 anos e de 10 e 11 anos, no estudo de Reynolds (2007) a amostra foi formada por participantes da amplitude total de idades para a qual o Koppitz-2 pode ser utilizado. Considerando que foram encontradas diferenças nos resultados para as idades de 6 e 7 e de 10 e 11 anos separadamente (somente as médias de 6 e 10 anos se diferenciaram significativamente entre o teste e o reteste para estes grupos etários) e que há evidências em diferentes pesquisas que a idade pode interferir significativamente no desempenho do teste (principalmente entre as crianças), pode-se levantar a hipótese de que esta variável tenha colaborado para as diferenças encontradas nos dois estudos.

Koppitz, em 1975, também mostra em uma tabela as correlações teste-reteste obtidas em nove estudos, que variaram em relação ao número e à série escolar dos participantes e ao intervalo de tempo entre as duas aplicações. Concluiu que o Teste de Bender de acordo com o seu sistema de avaliação é bastante estável, com correlações variando de 0,53 a 0,88. As correlações foram mais baixas para crianças com disfunção cerebral mínima e/ou com problemas emocionais. Segundo essa autora um progresso instável e inconsistente no teste reflete um funcionamento instável do examinando e não uma falta de precisão da avaliação do teste.

Rueda, Suehiro e Silva (2008) empreenderam uma pesquisa com 80 crianças de ambos os sexos com idades entre 6 e 11 anos ($M=8,90$; $DP=1,38$) que fizeram o Bender duas vezes com o intervalo de um mês entre as aplicações. Os testes foram aplicados coletivamente e avaliados por dois psicólogos, seguindo o modelo do Sistema de Pontuação Gradual (Sisto et al., 2005), pois o estudo também pretendia verificar a precisão entre avaliadores. Os autores apresentaram as correlações da precisão teste e reteste para o total de pontos e por figura. As correlações por figura para o Avaliador 1 variaram entre 0,27 e 0,69, para o Avaliador 2 entre 0,18 e 0,64 e para a pontuação 0,70 e 0,73 respectivamente para os Avaliadores 1 e 2. Os autores concluíram que, ainda que as correlações por figura tenham sido menores do que as esperadas em pesquisas de teste-reteste, podem ser consideradas satisfatórias, uma vez que no Bender o protocolo é avaliado como um todo e os resultados para o

total de pontos estão dentro dos parâmetros encontrados na literatura científica. Os resultados de Rueda et al. (2008) foram um pouco menores do que os do presente estudo e os de Reynolds (2007), no entanto, a possibilidade de comparação é um pouco limitada, uma vez que os métodos de aplicação, avaliação e o número de figuras são diferentes entre os sistemas. A opção de citar o estudo foi feita para mostrar a proximidade de coeficientes de técnicas semelhantes.

A fidedignidade por meio da consistência interna foi obtida com os dados da amostra total da pesquisa. O Alfa de Cronbach para 6 e 7 anos foi de 0,794 e para 8 a 12 anos de 0,843, sendo considerados adequados (Pasquali, 2003 e Hair et al., 2009), bem como confirmam a precisão do teste. Em relação às idades o alfa mais elevado foi para 12 anos (0,879), seguido pelo de 9 anos (0,847) e de 8 anos (0,819). O menor alfa foi o de 7 anos (0,716). A média do alfa para toda a amostra foi de 0,815. Pode-se verificar ainda que há um aumento dos alfas entre os 8 e 9 anos, um declínio aos 10 e novo aumento aos 11 e 12 anos, indicando certa tendência de aumento com o crescimento da idade. Reynolds (2007) obteve resultados muito semelhantes: 0,81 para 6 anos; 0,80 para 7 anos; 0,84 para 8 e 9 anos; 0,88 para 10 anos; 0,91 para 11 e 0,89 para 12 anos. No entanto, a tendência de aumento do alfa para as crianças mais velhas não foi observada na amostra americana. A média dos alfas também foi um pouco mais alta (0,853) em relação à brasileira. Os resultados por idades apresentados por este autor foram muito parecidos com o da pesquisa atual: 0,80 para crianças de 5 a 7 anos e 0,91 para os de 8 ou mais. O alfa do último grupo pode ter sofrido a influência da diferença na composição das amostras, já que os dados americanos incluíram os participantes mais velhos. Quanto ao gênero, nos dois grupos americanos o os meninos tiveram correlações maiores do que as das meninas, ainda que muito próximos (para o mais novos 0,79 para as meninas e 0,80 para os meninos e para os mais velhos 0,92 para os de sexo masculino e 0,91 para o feminino). Na pesquisa brasileira os resultados dos meninos foram maiores (0,793) do que os das meninas (0,785), para as crianças mais novas e para as mais velhas foram iguais (0,842).

A precisão entre avaliadores é outro tipo de medida de precisão associada ao erro de medida do teste, causado pela variabilidade na pontuação atribuída pelo avaliador, portanto está relacionada à compreensão e respeito aos critérios de avaliação. Koppitz (1975) apresenta para a sua escala as correlações de 23 estudos

a esse respeito entre 1965 e 1973, com 2 a 3 avaliadores que variaram de 0,79 a 0,99, das quais 81% delas eram maiores do que 0,89.

Esse tipo de precisão considerando a pontuação total, nesta pesquisa foi de 0,966 para 6 e 7 anos e 0,972 para os de 8 a 12 anos. Reynolds (2007) relata uma pesquisa em que 60 protocolos de pacientes com encaminhamento clínico (30 de 5 a 7 anos e 30 de 8 a 47 anos) foram avaliados por dois especialistas no Bender. As correlações foram 0,91 para 5 a 7 anos e 0,93 para 8 a 47 anos, muito próximos aos do presente estudo, sendo mais uma evidência para confirmar a precisão do Koppitz-2. Rueda et al. (2008), no estudo já citado, também calcularam a precisão entre avaliadores e obtiveram coeficientes semelhantes para o total de pontos (0,92 na primeira aplicação e 0,88 na segunda). As correlações por figura variaram entre 0,47 e 0,89 na primeira aplicação e de 0,46 e 0,81 na segunda.

Quanto à concordância entre os juízes para os itens do Koppitz-2 verificou-se que os coeficientes Kappa variaram entre 0,475 e 0,919 para os mais novos e entre 0,554 e 1,000 para os mais velhos, indicando uma concordância entre moderada e quase perfeita para todos os itens. Para os itens que não permitiram o cálculo do Kappa as frequências de pontos semelhantes foram muito altas. A metade dos itens dos mais novos apresentou concordância superior a 0,80 e para os mais velhos isso ocorreu em 35,6% dos itens. Somente sete itens, três das crianças mais novas e quatro das mais velhas apresentaram concordância moderada. Portanto, os resultados apontaram que tanto em relação ao total de pontos quanto por itens os coeficientes foram muito satisfatórios e podem ser considerados como evidência de fidedignidade satisfatória para o sistema Koppitz-2.

Para a análise dos itens foram calculadas as correlações item-total (r_{it}) para os dois grupos, que mostra quanto o item contribui para o total de pontos do teste. Além disso, permite verificar, se o item é discriminativo, ou seja, se ele contribui para discriminar os grupos em função da característica que está sendo avaliada no teste. Os parâmetros para considerar um item adequado variam. Guilford e Fruchter (1978) consideram que o valor mínimo aceitável é 0,30. Erthal (1987) estabelece o valor de 0,25, por outro lado Rueda e Muniz (2012) e Rueda e Castro (2012) afirmam que o *American Institute for Research* estabelece 0,20 como indicador de boa discriminação do item. Para a análise de itens do Koppitz-2 foi considerado o parâmetro de 0,20, como mínimo para a correlação item-total.

Os resultados desta pesquisa mostram que foram observadas evidências de validade, precisão e de qualidade dos itens satisfatórias que confirmam as propriedades psicométricas do Koppitz-2, ainda que seja necessário alguns ajustes nos critérios de avaliação para auxiliar no trabalho do psicólogo que utilizar o instrumento em seus diagnósticos.

VI. CONSIDERAÇÕES FINAIS

O Teste Gestáltico Visomotor de Bender é um dos testes mais tradicionais, que é utilizado em avaliações neuropsicológicas, psicológicas e psicodiagnósticos infantis, para as mais diversas finalidades, ainda que existam sistemas de avaliação que também possibilitam seu uso na avaliação de adolescentes e adultos. Como já foi mencionado anteriormente, existe na literatura uma grande variedade de propostas e metodologias de avaliação para diferentes finalidades (maturidade percepto-motora, avaliações cognitivas, de personalidade entre outras). No entanto, atualmente no Brasil somente o Teste Gestáltico Visomotor de Bender – Sistema de Pontual Gradual (B-SPG) pode ser utilizado para fins diagnósticos uma vez que é o único sistema que está aprovado pelo Conselho Federal de Psicologia (CFP), de acordo com a Resolução 002/2003 e que se destina a crianças com idades entre 6 e 10 anos. Os demais sistemas somente podem ser usados para fins de pesquisa.

Neste contexto tornou-se importante ampliar os estudos com o teste a fim de serem validados outros sistemas de avaliação. Como a proposta de Koppitz é a mais conhecida, utilizada e pesquisada seja no Brasil ou no exterior, a primeira ideia do pesquisador foi realizar novos estudos com a mesma. Foi quando se deparou com uma nova versão, o Koppitz-2, que constitui, não só uma revisão da proposta inicial, mas um redesenvolvimento com a ampliação de sua utilização. Segundo o autor, esta nova edição tem cinco usos principais: 1) constatar a presença e o grau de dificuldades visomotoras nos examinandos; 2) identificar candidatos a encaminhamentos para outros tratamentos e diagnósticos mais específicos; 3) verificar a efetividade de programas de intervenção desenvolvidos para corrigir problemas visomotores, que podem interferir não somente nos processos de aprendizagem, mas na vida de pessoa como um todo; 4) como instrumento de pesquisas e 5) colaborar em diagnósticos diferenciais de diversas condições psicológicas e neuropsicológicas ao longo da vida. A ênfase do sistema de pontuação do Koppitz-2 é avaliar a progressão do desenvolvimento quanto à capacidade de eliminar erros específicos na reprodução de figuras. A literatura consultada possibilitou também constatar a importância da capacidade de integração visomotora que pode interferir na vida das pessoas como um todo e não somente nos processos de aprendizagem.

Dessa forma, surgiu a proposta da presente pesquisa, sendo então elaborada uma metodologia, que tinha por finalidade verificar as propriedades psicométricas do Koppitz-2, o que levou à realização de diferentes estudos de validade e fidedignidade, além da normatização. Uma vez que a amplitude de idade do Koppitz-2 era muito grande, optou-se trabalhar apenas com crianças de 6 a 12 anos, período em que a maturidade percepto-motora se encontra em fase de desenvolvimento. De forma geral, pode-se dizer que foram encontrados coeficientes de validade e precisão de diferentes tipos e com diferentes grupos, que comprovam as suas qualidades psicométricas. Tendo sido encontradas evidências que comprovam a relação positiva entre a idade e a maturidade percepto-motora (integração visomotora, segundo Reynolds), a melhora do desempenho com o avanço da idade; relação com outro instrumento que avalia a mesma característica; com um teste de inteligência não verbal e com o sistema de avaliação original do teste. Em relação à fidedignidade foram obtidos coeficientes adequados de teste e reteste (estabilidade temporal), entre avaliadores e por consistência interna, que confirmaram a qualidade do instrumento, além da análise de itens, que também foi satisfatória.

Uma das dificuldades para a realização deste trabalho foi a quantidade reduzida de artigos publicados sobre o Koppitz-2 na literatura internacional, mesmo o material tendo sido publicado em 2007, e nenhum artigo no Brasil. Esta falta de literatura sobre o teste prejudicou um pouco a discussão, fazendo com que fosse necessário incluir estudos semelhantes aos que foram realizados aqui, mas com outros sistemas de avaliação, limitando as comparações aos estudos descritos no manual do teste. Apesar disso, foi possível verificar que muitos dos resultados desta pesquisa são semelhantes aos dos estudos realizados com sistemas já consolidados, como os de Koppitz, bem como sobre a relação do Bender com a inteligência e da possibilidade de avaliar o desenvolvimento da integração psicomotora. O diferencial deste novo sistema, em que foram incluídas novas figuras, com critérios de avaliação diferenciados em função da idade do examinando, mostra uma necessidade de que toda técnica ou instrumento de avaliação psicológica deveria apresentar: a atualização, não só no sentido de elaboração de novas normas de avaliação, que acompanhem a realidade socioeconômica e cultural da sociedade, mas principalmente buscando ampliar o uso e auxiliar, não só o psicólogo que trabalha com o teste, mas também o examinando que será avaliado, pois, em geral, e especificamente em relação ao Bender Koppitz-2 seu uso se

concentra principalmente na área clínica e de psicodiagnósticos para diversos contextos.

Como tudo que é novo precisa ser analisado e acaba passando por ajustes que têm a finalidade de melhorar o que foi desenvolvido, este trabalho também mostrou a necessidade de, por exemplo, modificar alguns dos critérios de avaliação. Certamente é preciso que sejam realizadas novas pesquisas com o Koppitz-2 para conhecer a opinião de outros pesquisadores em pesquisas com outros grupos, para que se possa aprimorar o instrumento, propondo ajustes nos critérios de avaliação.

É importante pontuar também algumas limitações da presente pesquisa. Entre elas, deve-se citar a necessidade de aumentar os grupos por faixa etária, incluindo principalmente participantes de escolas particulares, uma vez que a maioria de estudos com crianças considera esta questão, pois há evidências na literatura científica brasileira da sua influência nos resultados dos testes. Além disso, o aumento do tamanho da amostra também colabora para sua maior representatividade, ampliando a sua precisão. Em relação aos critérios de avaliação de alguns itens, ficou evidente a necessidade de revisões que forneçam definições mais detalhadas para aumentar sua objetividade.

Novos estudos também devem considerar a ampliação das faixas de idade, uma vez que o Koppitz-2 também pode ser administrado em adultos e idosos. Considerando que a expectativa de vida do brasileiro tem aumentado, cresce também a necessidade de instrumentos que possam auxiliar os profissionais nas avaliações para a população mais velha. Outra frente de estudo é a atualização da pesquisa com os indicadores emocionais do Koppitz-2, ponto este bastante discutido e controverso na literatura, em que os autores adotam posições bastante divergentes de apoio ou de questionamento para este uso, uma razão a mais para que novas pesquisas sejam realizadas com este foco.

Assim espera-se que os resultados desta pesquisa possam auxiliar os psicólogos na única atividade que é privativa de sua profissão, a avaliação psicológica.

VII. REFERÊNCIAS BIBLIOGRÁFICAS

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Andrés-Pueyo, A. (2006). Modelos psicométricos da inteligência. In: C. Flores-Mendoza, & R. Colom (Orgs.), *Introdução à Psicologia das diferenças individuais*. (pp. 73-100). Porto Alegre: Artmed.
- Allen, R. A., & Decker, S. L. (2008). Utility of The Bender Visual Motor Test – Second Edition in the assessment of attention-deficit / hyperactivity disorder. *Perceptual and Motor Skills*, 107(3), 663-675.
- Alves, I.C.B. (2002). *Teste de Inteligência Não Verbal R-1. Manual*. São Paulo: Vetor Editora.
- Alves, I. C. B., Alchieri, J. C., & Marques, K. (2002). As técnicas de exame psicológico ensinadas nos cursos de graduação de acordo com os professores. *Psico-USF*, 7(1), 77-88.
- Anastasi, A., & Urbina, S. (2000). *Testagem psicológica*. Porto Alegre: Artmed.
- Baltar, V. T., & Okano, V. (s.d.). *Análise de concordância – Kappa. Lee – Laboratório de Epidemiologia e Estatística*. Recuperado em 29 de maio de 2013 de <http://lee.dante.br/pesquisa/kappa/#author> .
- Bartholomeu, D., & Sisto, F. F. (2008). Maturidade viso-motora e inteligência: Um estudo correlacional. *Psicologia: Ciência e Profissão*, 28(2), 362-373.
- Beery, K. E. (1989). *Developmental Test of Visual-Motor Integration*. (3rd ed.) Toronto: Modern Curriculum Press.

- Beery, K. E., & Beery, N. A. (2004). *The Beery-Buktenica Developmental Test of Visual-Motor Integration (Beery VMI): Administration, scoring, and teaching manual* (5th ed.). Minneapolis, MN: NCS Pearson.
- Bender, L. (1955). *Test Gestaltico Visomotor (B-G): Uso y aplicaciones clinicas*. Buenos Aires: Paidós.
- Bender, L. (2011). *Test Gestaltico Visomotor (B-G): Uso y aplicaciones clinicas*. Buenos Aires: Paidós (Original publicado em 1955).
- Bernstein, J. (1955a). El Teste de Bender: Fundamentos científicos. In: L. Bender, *Test Gestaltico Visomotor*. (pp. 09-16). Buenos Aires: Paidós.
- Bernstein, J. (1955b). Introducción. La Psicología factorial de Spearman. In: C. Spearman. *Las habilidades del hombre, su naturaleza y medición*. (pp. 11- 21). Buenos Aires: Paidós.
- Bernstein, J. (2011). El Teste de Bender: Fundamentos científicos. In: L. Bender. *Test Gestaltico Visomotor*. (pp. 9-16). Buenos Aires: Paidós (Original publicado em 1955).
- Brannigan, G.G., & Brannigan, M. J. (1995). Comparison of individual versus group administration of the Modified Version of the Bender Gestal Test. *Perceptual and Motor Skills*, 80(3 pt 2), 1274.
- 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 Company.
- Brannigan, G. G., & Brunner, N. A. (1991). Relationship between two scoring systems for the Modified Version of The Bendet-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 on the Bender-Gestalt Test. *Journal of School Psychology, 31*, 327-330.
- Brannigan, G. G., & Brunner, N. A. (2002). *Guide to the qualitative scoring system for the Modified Version of the Bender-Gestalt Test*. Springfield, Illinois: Charles C Thomas Publisher, LTD.
- Brannigan, G. G., & Decker, S. L. (2003). *Bender Visual-Motor Gestalt Test, Second Edition*. Itasca, IL: Riverside Publishing.
- Brannigan, G. G., Decker, S. L. & Madsen, D. H. (2004). Innovative features of the *Bender-Gestalt II* and expanded guidelines for the use of the global scoring system. *Assessment Service Bulletin – Number 1*. Recuperado em 28 de maio de 2013 de http://www.riversidepublishing.com/products/bender/pdf/BenderII_ASB1.pdf
- Byrt, T. (1996). How good is that agreement? *Epidemiology, 7*(5), 561.
- Clawson, A. (1992). *Bender infantil: Manual de diagnóstico clínico*. Porto Alegre: Artes Médicas. (Original publicado em 1962).
- Cohen, R. J., Swerdlik, M. E., & Sturman, E. D. (2014). *Testagem e avaliação psicológica: Introdução a testes e medidas*. Porto Alegre: McGrawHill Education e Artmed.
- Conselho Federal de Psicologia. (2003). *Resolução CFP Nº 002/2003*. Recuperado em 30 de maio de 2013 de http://site.cfp.org.br/wp-content/uploads/2003/03/resolucao2003_02_Anexo.pdf
- Cunha, J. A. (2000a). Bender na criança e no adolescente. In: J. A. Cunha e cols. *Psicodiagnóstico V*. (5ª edição revisada e ampliada). (pp. 295-316). Porto Alegre: Artes Médicas.

- Cunha, J. A. (2000b). Bender no adulto. In: J. A. Cunha e Cols. *Psicodiagnóstico V*. (5ª edição revisada e ampliada). (pp. 317-339). Porto Alegre: Artes Médicas.
- Dancey, C.P., & Reidy, J. (2006). *Estatística sem Matemática para Psicologia usando SPSS para Windows*. Porto Alegre: Artes Médicas.
- Dancey, C.P., & Reidy, J. (2013). *Estatística sem Matemática para Psicologia*. Porto Alegre: Penso.
- De Filippi, G., & Kato, R. A. F. (2010). Teste Bender versão B-SPG. In: L. F. Malloy-Diniz, D. Fuentes, P. Mattos, & N. Abreu (Orgs.). *Avaliação Neuropsicológica*. (pp 362-366). Porto Alegre: Artmed.
- Domingues, S. F. S., Alves, I. C. B., Rosa, H. R., & Sargiani, R. A. (2012). As técnicas gráficas na avaliação cognitiva e da organização visomotora. In: S. M. Wechsler; T. C. Nakano. (Orgs.). *O desenho infantil: Forma de expressão cognitiva, criativa e emocional*. (pp 97-126). São Paulo: Casa do Psicólogo.
- Erthal, T. C. (1987). *Manual de Psicometria*. Rio de Janeiro: Zahar.
- Flores-Mendoza, C. E., Mansur-Alves, M., Abad, F. J., & Lelé, A. J. (2010). O que mede o Desenho da Figura Humana?: Estudos de validade convergente e discriminante. *Boletim de Psicologia*, 60(132), 73-84.
- Fonseca, W. C. (2003). *Padronização da Bateria de Orientação Profissional (BOP) para adolescentes de São Paulo*. Tese de Doutorado. Instituto de Psicologia Universidade de São Paulo, São Paulo.
- Gorske, T. T. (2007). Test reviews: Reynolds, C. R. (2007). Koppitz Developmental Scoring System for The Bender Gestalt Test (Koppitz-2). Austin, TX: Pro-ed. *Journal of Psychoeducational Assessment*, 26, 195-201.
- Guilford, J. P. (1950). *Fundamental statistics in Psychology and Education*. (2nd ed.). New York: McGraw-Hill.

- Guilford, J. P., & Fruchter, B. (1978). *Fundamental statistic in Psychology and Education*. (6th ed.). New York: McGraw-Hill.
- Hair Jr., J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). *Análise multivariada de dados*. Porto Alegre: Bookman.
- Hutt, M. (1969). *The Hutt adaptation of the Bender-Gestalt Test*. New York: Grune & Stratton.
- Hutt, M.L. (1975). *La adaptación Hutt del Test Guestaltico de Bender*. Buenos Aires : Editorial Guadalupe.
- Hutt, M.L. (1985). *The Hutt adaptation of the Bender-Gestalt test*. (4th ed.). Orlando: Grune & Stratton, Inc.
- Koppitz, E. M. (1974). *El Test Gestaltico Visomotor para niños*. Buenos Aires: Guadalupe. (Original publicado em 1963).
- Koppitz, E.M. (1975). *The Bender Gestalt Test for young children. Volume II: Research and application, 1963-1973*. New York: Grune & Stratton.
- Koppitz, E. M. (1989). *O Teste Guestáltico Bender para crianças*. Porto Alegre: Artes Médicas. (Original publicado em 1963).
- Kroeff, P. (1988). Normas brasileiras para o Teste de Bender. *Psicologia: Reflexão e Crítica*, 3(1/2), 12-19.
- Kroeff, P. (1991). Desempenho no Teste de Bender e nível sócio-econômico cultural. *Psicologia: Reflexão e Crítica*, 5(2), 119-126.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

- Marnic, L. R. (2011). Evaluating the Bender Visual Motor Gestalt Test II as a diagnostic screening instrument among clinically referred children and adolescents. *Dissertation Abstracts Internacional - Section B: The Sciences and Engineering*, 72(5-B), 3118.
- McDonald, C. A., Volke, M. A., Lopata, C., Toomey, J. A., Thomeer, M. L., Lee, G. K., Lipinski, A. M., Dua, E. H., Schivo, A. M., Bain, F., & Nelson, A. T. (2014). VMI-VI and BG-II Koppitz-2 for youth with HFASDs and typical youth. *Journal of Psychoeducational Assessment*, 32(5), 379-389.
- 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, 766.
- Nick, E., & Kellner, S. R. (1971). *Fundamentos de Estatística para as ciências do comportamento*. Rio de Janeiro: Renes.
- Noronha, A. P. P., Beraldo, F. N. M., & Oliveira, K. L. (2003). Instrumentos psicológicos mais conhecidos e utilizados por estudantes e profissionais de Psicologia. *Psicologia Escolar e Educacional*, 7(1), 47-56.
- Noronha, A. P. P., Primi, R., & Alchieri, J. C. (2005). Instrumentos de avaliação mais conhecidos/utilizados por psicólogos e estudantes de Psicologia. *Psicologia Reflexão e Crítica*, 18(3), 390-401.
- Nunes, C. H. S. S., & Primi, R. (2010). Aspectos técnicos e conceituais da ficha de avaliação dos testes psicológicos. Em Conselho Federal de Psicologia (Org.). *Avaliação psicológica: Diretrizes na regulamentação da profissão* (pp. 101-127). Brasília: CFP.
- Nunes, M. L. T., Ferreira, R. B., & Lopes, F. (2007). Diferentes sistemas de aplicação e interpretação do Teste Gestáltico Visomotor de Bender. *PSIC – Revista de Psicologia da Vetor Editora*, 8(1), 41-49.

- Oliveira, M. S. (1999). *Figuras Complexas de Rey: Teste de Cópia e Reprodução de Memória de Figuras Geométricas Complexas*. São Paulo: Casa do Psicólogo.
- Oliveira, M. S., & Rigoni, M. S. (2010). *Figuras Complexas de Rey: Teste de Cópia e Reprodução de Memória de Figuras Geométricas Complexas*. São Paulo: Casa do Psicólogo.
- Oliveira, R. (1973). *R-1: Teste Não Verbal de Inteligência – Manual*. São Paulo: Vetor Editora.
- Oliveira, R. (1997). *Manual do Teste R-2 (para crianças)*. São Paulo: Vetor Editora (edição experimental).
- Pasquali, L. (2003). *Psicometria: Teoria dos testes na Psicologia e na Educação*. Rio de Janeiro: Editora Vozes.
- Piotrowski, C. (1992). Foreword. In: N. Reichenberg & A. J. Raphael (1992). *Advanced psychodiagnostic interpretation of the Bender Gestalt Test: Adults and children*. (xiii – xvi). New York: Praeger Publishers.
- Raven, J. C. (1947). *Progressive Matrices. Sets A, Ab, B*. Londres: H.K. Lewis & Co.
- Reichenberg, N., & Raphael, A. J. (1992). *Advanced psychodiagnostic interpretation of the Bender Gestalt Test: Adults and children*. New York: Praeger Publishers.
- Reynolds, C. R. (2007). *Koppitz Developmental Scoring System for the Bender Gestalt Test - Second Edition (KOPPITZ-2)*. Austin, Texas: Pro-ed Inc.
- Reynolds, C. R., & Fletcher-Jansen, E. (2000). *Encyclopedia of special Education (2a. ed.)*. New York: Willey.
- Rosa, H. R., & Alves, I.C.B. (2012a). *Teste Não Verbal de Inteligência para Crianças – R-2*. São Paulo: Vetor Editora.

- Rosa, H. R., & Alves, I. C. B. (2012b). Relação do teste de Bender (avaliação Koppitz) com o R-2: Teste Não Verbal de Inteligência para Crianças. *Psicologia Teoria e Prática*, 14(1), 153-167.
- Rosa, H. R., & Alves, I.C.B. (2000). *Teste Não Verbal de Inteligência para Crianças – R-2*. São Paulo: Vetor Editora.
- Rueda, F. J. M., & Castro, N. R. (2012). *Teste de Inteligência (TI)*. São Paulo: Vetor Editora.
- Rueda, F. J. M., & Muniz, M. (2012). *Teste dos Cubos: Para avaliação do raciocínio visuoespacial*. São Paulo: Vetor Editora.
- Rueda, F. J. M., Bartholomeu, D., & Sisto, F. F. (2006). Maturidade perceptual e inteligência. *Psicologia: Ciência e Profissão*, 26(3), 490-503.
- Rueda, F. J. M., Suehiro, A. C. B., & Silva, M. A. (2008). Precisão entre avaliadores e pelo método do teste-reteste no Bender-Sistema de Pontuação Gradual. *Psicologia Teoria e Prática*, 10(1), 25-35.
- Santucci, H., & Galifret-Granjon, N. (1968). Prova Gráfica de Organização Perceptiva. In: R. Zazzo, *Manual para o exame psicológico da criança*. (pp. 232-268). São Paulo: Mestre Jou. (Original publicado em 1960).
- Santucci, H., & Pêcheux, M.G. (1981). Prova Gráfica de Organização Perceptiva para Crianças de 6 a 14 Anos. In: R. Zazzo, *Manual para o exame psicológico da criança*. (vol. 1, pp. 291-396). São Paulo: Mestre Jou. (Original publicado em 1969).
- Schelini, P. W. (2002). *Bateria Multidimensional de Inteligência Infantil: proposta de instrumento*. Tese de Doutorado. Pontifícia Universidade Católica de Campinas. Campinas, S. P.
- Siegel, S., & Castellan Jr., N. J. (2006). *Estatística não-paramétrica para ciências do comportamento*. Porto Alegre: Artmed.

- Silva, J. A. (2003). *Inteligência humana: Abordagens biológicas e cognitivas*. São Paulo: Lovise.
- Silva, R. B. F., & Nunes, M. L. T. (2007). Teste Gestáltico Visomotor Bender: Revendo sua história. *Avaliação Psicológica*, 6(1), 77-88.
- Sisto, F. F. (2005). *Desenho da Figura Humana DFH - Escala Sisto*. São Paulo: Vetor Editora.
- Sisto, F. F. (2007). Delineamento correlacional. In: M. N. Baptista & D. C. Campos. *Metodologia de pesquisa em ciências: Análises quantitativa e qualitativa*. Rio de Janeiro: LTC.
- Sisto, F. F., Bartholomeu, D., Rueda, F. J. M., Santos, A. A. A., & Noronha, A. P. P. (2008). Relações entre os testes de Bender e Matrizes Progressivas Coloridas de Raven na Avaliação da Inteligência. *Interação em Psicologia*, 12(1), 11-19.
- Sisto, F. F.; Noronha, A. P. P., & Santos, A. A. A. (2005). *Teste Gestáltico Visomotor de Bender: Sistema de Pontuação Gradual (B-SPG)*. São Paulo: Vetor.
- Soto, C. M. (2011). Test Gestáltico Visomotor de Bender Modificado y Test de Caras: una evaluación de la validez de constructo. *Cuadernos de Neuropsicología – Panamerican Journal of Neuropsychology*, 5(2), 129-143.
- Soto, C. M. (2013). Dimensionalidad del Test Gestáltico Visomotor de Bender Modificado: un análisis factorial confirmatorio. *Avaliação Psicológica*, 12(3), 341-350. Recuperado em 03 de maio de 2015, de http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1677-04712013000300009&lng=pt&tlng=es.

Spearman, C. (1955). *Las habilidades del hombre. Su naturaleza y medición*. Buenos Aires: Paidós. (Original publicado em 1927).

Urbina, S. (2007). *Fundamentos da testagem psicológica*. Porto Alegre: Artmed.

Volker, M. A., Lopata, C., Vujnovic, R. K., Smerbeck, A. M., Toomey, J. A., Rodgers, J. D., Schiavo, A., & Thomeer, M. L. (2010). Comparison of the Bender Gestalt-II and VMI-V in samples of typical children and children with high-functions autism spectrum disorders. *Journal of Psychoeducational Assessment*, 28(3), 187-200.

VIII. ANEXOS

ANEXO I

Termo de Consentimento Livre e Esclarecido

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (TCLE)

Estamos realizando uma pesquisa de Padronização Brasileira do Teste Guestáltico Visomotor de Bender – Koppitz - 2 para crianças e adolescentes entre 5 e 12 anos de idade. Trata-se de um teste psicológico que avalia a capacidade de integração visomotora, entendida como a capacidade percepção de figuras e a cópia adequada das mesmas. Para isso necessitamos de sua colaboração, que consiste em autorizar que seu filho(a) responda ao instrumento. A tarefa a ser realizada é a cópia de alguns cartões nos quais estão representados diversos desenhos (formas geométricas).

A aplicação deste teste não causa nenhum desconforto e/ou risco para o participante e somente será realizada com a sua concordância. A sua colaboração é muito importante, porém, *NÃO É OBRIGATÓRIA*. Caso não queira autorizar, sinta-se à vontade para recusar ou devolver a folha em branco.

Não há benefício direto para o participante, mas os resultados do estudo poderão auxiliar no trabalho dos psicólogos no diagnóstico psicológico.

As informações obtidas serão analisadas em conjunto com as de outras crianças, e não será divulgada a identificação de nenhum participante da pesquisa.

Não há despesas pessoais para o participante, nem compensação financeira relacionada à sua participação. Os pesquisadores se comprometem a utilizar os dados e o material coletado somente para esta pesquisa, cujos resultados grupais poderão ser publicados ou apresentados em Congressos Científicos.

Se estiver de acordo em autorizar a participação de seu filho(a) na pesquisa, solicitamos que preencha as informações a seguir e assine abaixo.

Eu, _____, portador(a) do documento de identidade _____, responsável pelo menor _____, consento de minha livre e espontânea vontade, na participação na pesquisa Padronização Brasileira do Teste de Bender Koppitz-2 para crianças a ser desenvolvido por Cristiano Esteves doutorando do Programa de Pós-Graduação do Departamento de Psicologia da Aprendizagem e do Desenvolvimento da Personalidade do Instituto de Psicologia da USP sob orientação da *Profa. Dra. Irai Cristina Boccato Alves*.

Sua assinatura neste termo de consentimento indica sua compreensão do objetivo e dos procedimentos do estudo e sua aceitação em participar do mesmo. Ainda assim, você pode desistir de participar dessa pesquisa a qualquer momento. Este Termo de Consentimento é feito em duas vias, sendo que uma permanecerá em seu poder e outra com o pesquisador responsável.

Agradecemos a colaboração e nos colocamos à disposição para maiores esclarecimentos.

São Paulo, ___ de _____ de 2014.

Assinatura do Responsável

ANEXO II

Instruções de Aplicação do Teste Bender Koppitz-2

ANEXO III

Folha de Registro do Examinador – 5 a 7 anos
Teste Bender Koppitz-2

FOLHA DE REGISTRO DO EXAMINADOR

TESTE DE BENDER – KOPPITZ II (5 a 7 anos)

Dados do Examinando:

Nome/ Iniciais : _____

Data de aplicação: ___/___/___ Sexo: () M () F Série/Ano: _____

Data de nascimento ___/___/___ Dominância Lateral: () Direita () Esquerda

Idade: ___anos ___ meses Aplicador: _____

Nome da Escola _____

Tipo de Escola: () Estadual () Municipal () Federal () Particular

Início da aplicação: _____ Término: _____ Tempo total: _____

Figura	Ordem e Direção	Observações	Itens	Pontos
1			1	
2			2	
3			3 4 5	
4			6	
5			7 8 9 10	

Figura	Ordem e Direção	Observações	Itens	Pontos
6			11	
7			12 13 14	
8			15 16	
9			17 18 19	
10			20 21 22	
11			23 24 25	
12			26 27 28 29 30	
13			31 32 33 34	

ANEXO IV

**Folha de Registro do Examinador – 8 a 85 anos ou mais
Teste Bender Koppitz-2**

FOLHA DE REGISTRO DO EXAMINADOR

TESTE DE BENDER – KOPPITZ II (8 a 85 anos ou mais)

Dados do Examinando:

Nome/ Iniciais : _____

Data de aplicação: ___/___/___ Sexo: () M () F Série/Ano: _____

Data de nascimento ___/___/___ Dominância Lateral: () Direita () Esquerda

Idade: ___anos ___ meses Aplicador: _____

Nome da Escola _____

Tipo de Escola: () Estadual () Municipal () Federal () Particular

Início da aplicação: _____ Término: _____ Tempo total: _____

Figura	Ordem e Direção	Observações	Itens	Pontos
5			1 2 3	
6			4	
7			5 6 7 8	
8			9 10	
9			11 12 13	

Figura	Ordem e Direção	Observações	Itens	Pontos
10			14 15 16 17	
11			18 19 20	
12			21 22 23 24 25 26	
13			27 28 29 30	
14			31 32 33 34 35	
15			36 37 38 39	
16			40 41 42 43 44 45	