

ALINE CHRISTINA ISTOME



ESTUDO PSICOMÉTRICO DA BATERIA INFORMATIZADA DE LINGUAGEM
ORAL (BILOV3): CORRELAÇÃO COM VOCABULÁRIO E INTELIGÊNCIA

Autora: Aline Christina Istome

Orientadora: Prof^a. Dr^a. Maria Cristina R. A. Joly

ITATIBA

2013

ALINE CHRISTINA ISTOME

ESTUDO PSICOMÉTRICO DA BATERIA INFORMATIZADA DE LINGUAGEM
ORAL (BILOV3): CORRELAÇÃO COM VOCABULÁRIO E INTELIGÊNCIA

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

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

ITATIBA

2013

153.99 Istome, Aline Christina.
I84b Bateria informatizada de linguagem oral (BILOv3)
correlação com vocabulário e inteligência. / Aline Christina
Istome. -- Itatiba, 2013.
117 p.

Dissertação (mestrado) – Programa de Pós-Graduação
Stricto Sensu em Psicologia da Universidade São Francisco.
Orientação de: Maria Cristina Rodrigues Azevedo Joly.

1. Teste informatizado. 2. Compreensão da linguagem oral.
3. Psicometria. 4. Avaliação psicoeducacional. I. Joly, Maria
Cristina Rodrigues Azevedo II. Título.

Dedicatória

Aos profissionais competentes que *“utilizam os testes de forma apropriada, profissional e ética, dando a devida atenção às necessidades e direitos de todas as partes envolvidas no processo de avaliação, bem como às razões que determinam a aplicação do teste e ao contexto no qual se realiza a testagem”*

International Test Commission (2000).

Agradecimentos

Ao fim desta etapa acadêmica, muito há que se agradecer. Juntamente com o Mestrado também ocorria o início da minha vida profissional no Lar Itatibense da Criança e o começo da experiência de ser mãe. Ter findado esta etapa com sucesso e total satisfação nas três áreas - acadêmica, profissional e pessoal - só foi possível porque tenho em minha vida pessoas especiais, que me deram todo o apoio e suporte para que pudesse chegar até aqui e que fizeram com que este momento fosse tão emocionante. Muito obrigada:

Aos meus pais, Maria José e Jefferson, e aos meus avôs e avós, por terem participado da minha formação humana e por todos os significados que introduziram na minha vida, sempre me apoiando, muitas vezes mesmo sem concordar com minhas escolhas;

Aos meus amores, meu marido Edgar, sempre companheiro, um grande parceiro de vida, e à nossa filha Laura, um presente, o maior e mais importante de todos os projetos;

Aos amigos, Jaque, Léia e James, por terem me permitido dividir sonhos e ideais, e por me ensinarem as lições mais preciosas e significativas sobre as relações humanas;

Aos parceiros, Diego, Nay, Ane, Luana, Samanta e Taís do Núcleo de Avaliação Psicológica Informatizada (NAPI) da Universidade São Francisco e do coração, pela generosidade e dedicação;

Um agradecimento especial ao meu amigo Diego, que foi um anjo da guarda, sempre presente, sempre um presente!;

Às minhas amigas do Lar Itatibense da Criança, Cláudia, Tomires, Fernanda, Marilandes, Rosângela e Roseli que permitiram que eu vivenciasse a prática do fazer Psicologia com ética, responsabilidade social e com muito prazer;

Aos professores da banca de defesa, Prof^a Claudette e Prof^o Ronei, que fizeram parte da minha história acadêmica desde o início e que tornaram os momentos de avaliação, momentos de compartilhamento de conhecimento, com preciosos apontamentos e sugestões;

À minha orientadora Maria Cristina Rodrigues Azevedo Joly que me acolheu desde o ingresso na graduação, acreditando em mim e me apoiando como uma mãe que deseja que o filho se desenvolva. Se não fosse por você eu não teria vivenciado esta grande conquista!

À Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP, pelo apoio financeiro.

Resumo

Istome, A. C. (2013). *Estudo psicométrico da Bateria Informatizada de Linguagem Oral (BILOv3): correlação com vocabulário e inteligência*. Dissertação de Mestrado, Programa de Pós-Graduação *Stricto-Sensu* em Psicologia, Universidade São Francisco, Itatiba.

A avaliação e detecção precoces de distúrbios na linguagem oral permitem a introdução eficaz de procedimentos de intervenção, diminuindo a incidência ou a severidade de problemas na alfabetização. Como conseguinte, considera-se importante o desenvolvimento de instrumentos com validade, precisão e normatização para avaliar a linguagem oral em crianças do Ensino Infantil e séries iniciais do Ensino Fundamental. Ao lado disso, considera-se que os testes informatizados incrementam as técnicas avaliativas, porquanto permitem a mensuração e testagem das habilidades cognitivas que não seriam possíveis por meio dos testes do tipo papel-e-lápis. Deste modo, o presente estudo objetivou avaliar a compreensão da linguagem oral receptiva, buscando evidências de validade para a Bateria Informatizada de Linguagem Oral versão 3 (BILOv3) por meio da correlação com o desempenho em vocabulário, avaliado pelo Teste de Vocabulário por Imagens Peabody informatizado (TVIP) e com o raciocínio, avaliado pelas Matrizes Progressivas Coloridas de Raven. Participaram do estudo 262 alunos com idades que variavam de cinco a oito anos ($M = 6,37$ e $DP = 0,89$), matriculados no último ano da Educação Infantil e no 1º e 2º ano do Ensino fundamental de escolas públicas do interior do Estado de São Paulo. Destes, 56,49% eram meninos e 43,51% meninas. A aplicação dos testes ocorreu em três sessões, uma para cada instrumento, em dias distintos, sendo a BILOv3 e o TVIP aplicados coletivamente na sala de informática e o Raven individualmente na sala de aula. Os resultados indicaram correlação positiva moderada entre o total da BILOv3 e o TVIP e entre o total da BILOv3 e o Raven, constatando-se a evidência de validade com base na relação com medidas convergentes. Além disso, a BILOv3 foi sensível para diferenciar os grupos extremos, formados pelos alunos com menores e alunos com maiores pontuações em vocabulário e no escore de inteligência. Foram encontradas diferenças estatisticamente significativas nos escores e no tempo de execução da BILOv3 em função da idade e da escolaridade, indicando evidências de validade com base na relação com critérios desenvolvimentais. Foi constatada a diferença significativa nas médias das provas SL, IH e no TVIP em relação ao sexo do aluno, indicando que as meninas obtiveram médias maiores do que a dos meninos. Foi observado efeito teto nas provas CM, IH, CF e CH em contraste com as baixas pontuações da SL e OLV que podem estar prejudicando a aferição da linguagem oral por meio da BILOv3. Por isso, considera-se que, com as alterações indicadas para estas provas, a BILOv3 poderá ser utilizada como instrumento eficiente na avaliação e detecção precoce de distúrbios na linguagem receptiva.

Palavras-chave: Teste informatizado, compreensão da linguagem oral, psicometria, avaliação psicoeducacional.

Abstract

Istome, A. C. (2013). *Estudo psicométrico da Bateria Informatizada de Linguagem Oral (BILOv3): correlação com vocabulário e inteligência*. Dissertação de Mestrado, Programa de Pós-Graduação *Stricto-Sensu* em Psicologia, Universidade São Francisco, Itatiba.

The precocious evaluation and detection of oral language disorders possibility efficient interventions procedures minimize the literacy problems incidence or the severity. As a result, It's important the development of tests with validity, reliability and norms to evaluate the oral language in preschool children and in the initial of elementary school. Besides this, the computerized tests increase the evaluative techniques because possibility the cognitive skills measurement and test that wouldn't be possible with paper-and-pencil tests. Thus, these study aimed to evaluate the oral receptive language comprehension to find validity evidences to Oral Language Computerized Test (Bilov3) by correlating with the vocabulary achievement, evaluated by the Computerized Peabody Picture Vocabulary Test (TVIP) and with the reasoning, evaluate with Raven Colored Progressive Matrices. The subjects were 262 students, from these 56,4% were boys, with age from Five to eight years old ($M = 6,37$ e $DP = 0,89$), enrolled in the last year of public preschool and in the first and second stage of elementary school. The testing was in three different sessions, one to each test used, in different days. The testing of BILOv3 and the TVIP was collectively in the informatics lab and the Raven was testing with each child in one classroom. The data showed the positive and moderate correlation between the BILOv3 total score with TVIP and Raven, that indicated the BILOv3 validity evidence based in the relation with convergent measures. Furthermore, the BILOv3 differed extremes groups organized by the Best and worse vocabulary and intelligence scores. There were statistically significant differences in the scores and execution time of BILOv3 as a function of age and schooling, showing validity evidences of external variable criteria. It was found average statistically significant differences of the Logical Sequence (SL), Story Interpretation (IH) and in TVIP in relation to student genre that indicates the girls best score as the boys. It was observed ceiling effect in Morphosintactica Comprehension (CM), IH, Phrases Comprehension (CF) and Story Comprehension (CH) in contrast to SL and Verbal and Logical Organization (OLV) that may be hindering the measurement of comprehension in oral language. Therefore, it is believed that, with the changes indicated for these proves, the BILOv3 may be used as a tool in assessing accountable, efficient and early detection of receptive language disorders.

Keywords: computerized test; oral comprehension language; psychometrics; psychoeducational evaluation

Sumário

LISTA DE FIGURAS	viii
LISTA DE TABELAS.....	x
APRESENTAÇÃO	1
1. LINGUAGEM.....	4
<i>1.1 CONCEITUAÇÃO E DESENVOLVIMENTO</i>	<i>4</i>
<i>1.2 AVALIAÇÃO DA LINGUAGEM.....</i>	<i>9</i>
<i>1.3 CARACTERÍSTICAS DOS INSTRUMENTOS DE AVALIAÇÃO</i>	<i>16</i>
<i>1.4 ESTUDOS COM A BATERIA INFORMATIZADA DE LINGUAGEM ORAL - BILO</i>	<i>23</i>
<i>1.5 ESTUDOS COM CONSTRUTOS CORRELATOS</i>	<i>36</i>
2. OBJETIVO	42
3. MÉTODO	43
<i>3.1 PARTICIPANTES</i>	<i>43</i>
<i>3.2. INSTRUMENTOS E MATERIAIS</i>	<i>45</i>
EXEMPLO:	55
<i>3.3. EQUIPAMENTOS.....</i>	<i>55</i>
<i>3.4. PROCEDIMENTOS.....</i>	<i>55</i>
4. RESULTADOS E DISCUSSÃO	59
<i>4.1 TESTE DE VOCABULÁRIO POR IMAGENS PEABODY – (TVIP).....</i>	<i>59</i>
<i>4.2 DESEMPENHO NAS MATRIZES PROGRESSIVAS COLORIDAS DE RAVEN</i>	<i>66</i>
<i>4.3 DESEMPENHO NA BATERIA INFORMATIZADA DE LINGUAGEM ORAL, BILOV3</i>	<i>75</i>
<i>4. 4 ANÁLISES CORRELACIONAIS</i>	<i>95</i>
5. CONSIDERAÇÕES FINAIS.....	106
5. REFERÊNCIAS.....	111

LISTA DE FIGURAS

Figura 1. <i>Modelo neuropsicolinguístico – MNPL (Chevrie-Muller & Narbona, 2005, p.104)</i>	12
Figura 2. <i>Distribuição do número de alunos por sexo em cada ano escolar frequentado...</i>	44
Figura 3. <i>Distribuição da idade dos alunos por nível escolar frequentado.</i>	45
Figura 4. <i>Exemplo 2 prova CM.</i>	47
Figura 5. <i>Exemplo 1 prova SL.</i>	48
Figura 6. <i>Exemplo 1 prova OLV.</i>	48
Figura 7. <i>Item 1 prova IH.</i>	49
Figura 8. <i>Item 2 prova CF.</i>	50
Figura 9. <i>Itens 1, 2 e 3 prova CH.</i>	51
Figura 10. <i>Exemplo de item do TVIP.</i>	53
Figura 11. <i>Exemplo de item Matrizes Progressivas Coloridas de Raven</i>	55
Figura 12. <i>Distribuição de frequência das pontuações dos alunos em vocabulário receptivo.</i>	61
Figura 13. <i>Diagrama de caixa com a distribuição da pontuação em vocabulário receptivo.</i>	62
Figura 14. <i>Distribuição de frequência das pontuações dos alunos em inteligência.</i>	68
Figura 15. <i>Diagrama de Caixa com a distribuição de frequência das pontuações dos alunos em inteligência.</i>	69

Figura 16. <i>Distribuição de frequência da classificação dos alunos segundo o nível de inteligência, por ano escolar.</i>	71
Figura 17. <i>Distribuição da pontuação em compreensão da linguagem oral.</i>	76
Figura 18. <i>Diagrama de caixa com a distribuição da pontuação em compreensão da Linguagem Oral.</i>	77

LISTA DE TABELAS

Tabela 1 <i>Estatísticas descritivas do desempenho dos alunos no Teste de Vocabulário por Imagens – Peabody - TVIP</i>	60
Tabela 2 <i>Distribuição de médias de vocabulário receptivo e do tempo de execução em função do ano escolar que os alunos estavam matriculados</i>	63
Tabela 3 <i>Distribuição de médias de vocabulário receptivo e do tempo de execução em função da idade dos alunos</i>	64
Tabela 4 <i>Distribuição de médias de vocabulário receptivo e do tempo de execução segundo o sexo dos alunos</i>	66
Tabela 5 <i>Estatísticas descritivas do desempenho dos alunos no teste Matrizes Progressivas Coloridas de Raven</i>	67
Tabela 6 <i>Distribuição de médias de inteligência e do tempo de execução em função do ano escolar que os alunos estavam matriculados</i>	70
Tabela 7 <i>Distribuição de médias de inteligência e do tempo de execução no Raven em função da idade dos alunos</i>	72
Tabela 8 <i>Distribuição de médias de inteligência e do tempo de execução do Raven segundo o sexo dos alunos</i>	73
Tabela 9 <i>Estatísticas da análise exploratória do desempenho dos alunos na BILOv3</i>	75
Tabela 10 <i>Estatísticas descritivas do desempenho nas provas da BILOv3</i>	78
Tabela 11 <i>Estatísticas descritivas do desempenho nas provas da BILOv3 nos cinco estudos anteriores e na presente pesquisa</i>	80

Tabela 12 <i>Estatísticas descritivas dos escores por prova e total da BILOv3 em função do ano escolar.</i>	81
Tabela 13 <i>Estatísticas descritivas dos escores por prova e total da BILOv3 em função da idade.</i>	84
Tabela 14 <i>Estatísticas descritivas dos escores por prova e total da BILOv3 segundo o sexo do aluno</i>	85
Tabela 15 <i>Estatísticas descritivas do tempo na realização da BILO, TVIP e Raven.</i>	88
Tabela 16 <i>Estatísticas descritivas dos tempos em segundos por prova e total da BILOv3 em função do ano escolar.</i>	90
Tabela 17 <i>Estatísticas descritivas dos tempos por prova e total da BILOv3 em função da idade.</i>	92
Tabela 18 <i>Estatísticas descritivas dos tempos em segundos por prova e total da BILOv3 segundo o sexo do aluno</i>	94
Tabela 19 <i>Correlação entre os escores da BILOv3, o Teste de Vocabulário por Imagens Peabody – TVIP e as Matrizes Progressivas Coloridas de Raven</i>	95
Tabela 20 <i>Coefficientes de correlação de Pearson entre a BILO e o TVIP em estudos anteriores</i>	96
Tabela 21 <i>Correlação total e parcial entre os escores da BILOv3 e do Teste de Vocabulário por Imagens Peabody – TVIP</i>	100
Tabela 22 <i>Correlação total e parcial entre os escores da BILOv3 e das Matrizes Progressivas Coloridas de Raven</i>	102

Tabela 23 <i>Estatísticas descritivas do desempenho na BILOv3 em função dos grupos extremos formados pelo TVIP</i>	103
Tabela 24 <i>Estatísticas descritivas do desempenho na BILOv3 em função dos grupos extremos formados com base nos escores de inteligência.</i>	105

APRESENTAÇÃO

A presente dissertação refere-se à busca de evidências de validade para um instrumento de avaliação da linguagem oral, a Bateria Informatizada de Linguagem Oral – BILO, destinada a avaliar crianças pré-escolares e de séries iniciais do Ensino Fundamental. Este trabalho vem em um *continuum* das pesquisas que vêm sendo desenvolvidas desde 2006, ano de início da construção da BILO. A importância do investimento neste instrumento pauta-se na premissa de que a avaliação da linguagem oral em crianças no início da escolarização mostra-se útil e imprescindível para a detecção precoce de dificuldades na aquisição da linguagem, que se concretizarão em futuras dificuldades na leitura e escrita (Capovilla, Joly, & Tonelotto, 2006, Pedrosa & Rotta, 2006).

A BILO foi construída tendo como referencial teórico o modelo psicolinguístico, que considera a linguagem como um instrumento ligado à cognição e à comunicação humana, considerando tanto a organização das palavras em sua função cognitiva, como também os efeitos do meio externo, que englobam a intenção de comunicação, o prazer da troca e do diálogo (Le Normand, 2005). Diante desta concepção, a BILO está pautada no exame do que a criança compreende a partir do que ouve, considerando o processamento da entrada da informação nos seus diferentes módulos de tratamento da informação. O modelo psicolinguístico adotado para nortear os testes da BILO foi o Modelo Neuropsicolinguístico, o MNPL, desenvolvido por Chrevie-Muller e Narbona em 2005. Este modelo permite a descrição dos diferentes níveis de funcionamento da linguagem, os aspectos formais, pragmáticos e anátomo-funcionais, permitindo identificar a patologia da linguagem na criança.

A BILO se encontra na terceira versão, sendo que os estudos anteriores para a versão1 (Joly, Martins, Zani, Istome, & Santos, 2008; Almeida & Joly, 2008; Issa, 2008; Joly & Piovesan, 2012; Joly & Dias, 2009) e para versão2 (Joly, Martins, Lopes, & Lemos, 2009; Joly, Piovezan, Soares, Lopes, & Martins, 2009) registram os estágios de validação no processo de construção e desenvolvimento, pelo qual o instrumento passou até a consolidação da versão atual (BILOv3). Nas duas versões anteriores, apesar de serem encontradas evidências de validade, foram indicadas alterações relativas às provas que compõem a bateria, sendo necessária a reformulação que constituiu a terceira versão.

Assim, o presente estudo tem como objetivo avaliar a compreensão da linguagem oral em alunos do último ano do Ensino Infantil e primeiro ao quinto ano do Ensino Fundamental de escolas de São Paulo, buscando evidências de validade pela relação com outras variáveis para a BILOv3, por meio do estudo correlacional com o desempenho em vocabulário e com a inteligência, construtos correlatos. Também serão verificadas diferenças relativas à série e a idade, que são critérios preditivos de desenvolvimento. Além disso, serão exploradas as possíveis diferenças entre o gênero dos participantes.

A correlação proposta para esta busca por evidências de validade está pautada na relação que o vocabulário, ou seja, o conhecimento do sentido da palavra tem com a compreensão da linguagem oral, bem como estas duas funções cognitivas estão ligadas diretamente ao desenvolvimento cognitivo como um todo, que pode ser mensurado por um escore de inteligência geral. Isto porque a compreensão da linguagem é dependente da interação dos processos de sensação, percepção, atenção, memória (de curto e longo prazo) e de funções executivas que subjazem e permitem o processamento de informações (Gerber, 1996).

Nesse sentido, pode-se afirmar que o fator geral de inteligência, ligado à capacidade educativa de relacionar (Spearman, 1927), esteja implicado na execução das provas de

linguagem, correlacionando-se de modo convergente. É comprovado que índices de vocabulário correlacionam-se positivamente com o quociente de inteligência (Strauss, Sherman, & Spreen, 2006). De acordo com Eysenck (1994), isto ocorre porque o uso de informações contextuais e frasais para inferir significado de palavras desconhecidas é um dos principais meios de adquirir vocabulário.

Considerando o objetivo, são abordados no Capítulo I os aspectos teóricos referentes à conceituação, ao desenvolvimento, à avaliação da linguagem, as características dos instrumentos de avaliação, assim como os estudos anteriores com a Bateria Informatizada de Linguagem Oral em suas diferentes versões e os construtos correlatos. No Capítulo II consta o detalhamento das questões de investigação. O método, descrito no Capítulo III, apresenta a descrição dos participantes, dos instrumentos e equipamentos utilizados, bem como o procedimento de aplicação e de correção dos testes. No Capítulo IV são apresentados e discutidos os resultados encontrados; o Capítulo V traz as considerações finais, e por fim, no Capítulo VI são apresentadas as referências utilizadas.

1. LINGUAGEM

1.1 Conceituação e desenvolvimento

A linguagem é um instrumento da comunicação interpessoal e do pensamento, sendo formada por um sistema estruturado de signos e códigos que exprimem ideias, as quais são manifestadas pela fala, gestos, escrita, incluindo o braile. Deste modo, a linguagem implica que um emissor codifique o significado de suas intenções em um código – sons, gestos ou escrita - e que um receptor decodifique o código em significados. Embora a linguagem seja um sistema de princípios e regras finito, ela possui a propriedade de ser infinitamente criativa, no que se refere à possibilidade de construção e entendimento de sentenças gramaticais novas (Gil, 2002; Kaufman, 1996). Assim, é considerada uma função superior do cérebro e, portanto, um ato complexo que envolve a cognição (Gerber, 1996).

Le Normand (2005) considera que o desenvolvimento da linguagem pode ser explicado a partir de três modelos, quais sejam, o Linguístico, o Cognitivo e o Psicolinguístico. O modelo Linguístico, considera a linguagem como inata, ou seja, geneticamente determinada, desenvolvida a partir da maturação neuronal, valorizando assim a predisposição biológica para a aquisição da linguagem. O modelo Cognitivo privilegia a noção de estratégia, isto é, como a criança consegue controlar as funções linguísticas (prosódica, fonológica, semântica, morfossintática e pragmática). Este modelo é abordado pela psicologia cognitiva que, utilizando-se das descrições linguísticas, caracteriza os esquemas das representações cognitivas envolvidas na linguagem, com a valorização das unidades de tratamento da informação. O Modelo psicolinguístico, surgiu considerando o modelo linguístico e o cognitivo, concebendo, desta forma, a linguagem

como um instrumento ligado à cognição e à comunicação humana, considerando tanto a organização das palavras em sua função cognitiva, como também os efeitos do meio externo, que englobam a intenção de comunicação, o prazer da troca e do diálogo (Le Normand, 2005).

Do ponto de vista psicolinguístico, é possível descrever o desenvolvimento das etapas da aquisição da linguagem em seus componentes fonológico, lexical, morfossintático, pragmático e metalinguístico. O desenvolvimento do “sistema fonológico”, que se refere ao inventário de sons e das regras para combiná-los em unidades significativas, pode ser observado na criança quando ela, no início da emissão de palavras, reproduz palavras ouvidas com modificações sistemáticas, caracterizadas por duplicação, omissão ou troca da posição das sílabas, além das trocas de vogais por consoantes, entre outras (Le Normand, 2005).

A partir dos 10 aos 13 meses a criança aprende o léxico, na razão de uma palavra por vez, configurando a etapa do desenvolvimento denominada de “sistema lexical”, que pode ser observado quando as crianças começam a emitir as primeiras palavras com significados. Observa-se que, no início, as primeiras palavras têm como função designar, expressar e dar ordens, de modo que a mesma palavra é generalizada para uma variedade de objetos e situações. A base lexical é constituída por um sistema que codifica os objetos familiares concretos, sendo que, geralmente as primeiras palavras são substantivos (Le Normand, 2005). Aos 12 meses a criança é capaz de compreender de 50 a 100 palavras, contudo, até um ano e meio de idade é capaz de utilizar em sua fala somente de 10 a 50 palavras (Pedroso & Rotta, 2006). Observa-se, assim, que seu vocabulário se amplia, sendo que a compreensão da palavra ouvida (léxico receptivo) precede a emissão das palavras (léxico expressivo). Cabe destacar a importância do vocabulário para a compreensão da linguagem que, conforme observado, é adquirido desde o início do desenvolvimento na

interação com o estímulo ambiental e com o exercício adequado das funções cognitivas, estando, portanto, diretamente imbricado na capacidade de compreensão da linguagem.

Já o desenvolvimento do “sistema morfossintático”, entre 2 e 3 anos, que consiste na organização das palavras segundo sua função na frase, inicialmente ocorre pela reprodução das sequências que os adultos utilizam em suas próprias produções. Posteriormente, aparece a “frase gramatical” cujos elementos fundamentais são a entonação (prosódia), o princípio de supergeneralização (generaliza uma conjugação verbal ou uma regra para todas as frases, ou ainda cria novas palavras por dedução sobre outras existentes), a flexão dos verbos e a ordem das palavras na frase. Nota-se que, neste período, as crianças passam a produzir de 150 a 200 palavras, formando frases constituídas por até três palavras (Le Normand, 2005; Pedroso & Rotta, 2006). A partir dos três anos a criança consegue se comunicar por frases gramaticais formadas por substantivos, verbo, adjetivo, artigos e preposição, além de conseguir usar o plural e formular questões. Aos quatro anos espera-se que a fala seja clara e coerente (Pedroso & Rotta, 2006).

O desenvolvimento das “capacidades pragmáticas”, que se referem à capacidade de compreender as intenções de comunicação e adequar o tipo de linguagem indicada a cada situação ou interlocutor, ocorre entre 6 e 7 anos. Por fim, ocorre o desenvolvimento das “capacidades metalinguísticas”, isto é, o domínio sobre o uso da língua, que vai permitir o discernimento de ambiguidades, a diferenciação de frases gramaticais das não-gramaticais e a manipulação da linguagem. Essa conscientização metalinguística se manifesta de modo heterocrônico entre os diferentes componentes do sistema linguístico fonético, semântico, morfossintático e pragmático (Le Normand, 2005). Pedroso e Rotta (2006) acrescentam que, enquanto o desenvolvimento fonêmico está finalizado aos 7 anos, a semântica é um processo longo que se desenvolve ao longo da vida.

Para Kaufman (1996), a competência linguística compreende o domínio sobre as “regras fonológicas”; sobre as “regras morfológicas”; sobre as “regras semânticas”; e sobre as “regras sintáticas”. Já a competência comunicativa compreende tanto a competência linguística quanto a percepção das regras que governam o uso da linguagem em contextos sociais, quais sejam elas: conhecimento de tipos de sentenças que são mais adequados para transmitir a mensagem; percepção da informação de fundo; e entendimento dos princípios cooperativos subjacentes às trocas na conversação, ou, como definidas por Le Normand (2005), as capacidades pragmáticas e metalinguísticas. Segundo Kaufman (1996), equipado com a competência linguística, o emissor e/ou receptor está livre para usar a linguagem no pensamento e na comunicação, sendo capaz de produzir e compreender uma variedade de tipos de sentenças, cada qual expressando um sentido diferente.

Diante do exposto, cabe considerar também que a linguagem, enquanto capacidade cognitiva, é dependente da interação dos processos de percepção, atenção e memória - de trabalho e de longo prazo- e da capacidade de controlar as respostas, pois estes subjazem o processamento de informações (Narbona & Fernandes, 2005). No modelo de processamento de informações cognitivo descrito por Andre e Phye (1986, citado por Gerber, 1996) são componentes da linguagem: o *input* ou registros sensoriais (sensação e percepção); memória de curto prazo (armazenamento breve para processamento geral de informações); memória de longo prazo (armazenamento de informações episódicas e semânticas); componente executivo (controle de informações, seleção e desempenho de operações); e o *buffer* de saída (organização da resposta ou *output* que engaja habilidades cognitivas aprendidas).

Observa-se então que a organização da linguagem se distribui em dois pólos, quais sejam, o pólo receptivo (*input* - porta de entrada que comporta não só a audição e a compreensão da linguagem falada, como, também, a visão e a compreensão da linguagem

escrita); e o pólo expressivo (*output* - porta de saída que comporta não só a fonação ou articulação verbal, como também a escrita) (Gil, 2002) . Assim, considera-se que dificuldades para compreender o código da linguagem falada indicam problema no pólo receptivo; e as dificuldades em aplicar as regras da linguagem para comunicar ideias e pensamentos indicam comprometimento do pólo expressivo (*American Speech-Language-Hearing Association - ASHA, 2007*).

Sobre estes aspectos, segundo Pedroso e Rotta (2006), os transtornos da linguagem são problemas comuns na infância, com uma prevalência estimada entre 1 e 12%, sendo que nas crianças pré-escolares a média é de 5%, com maior incidência sobre os meninos (de 2 a 4 para cada menina). Ainda segundo os autores, estes problemas de linguagem manifestados nas crianças com menos de cinco anos revelarão, em 60% dos casos, algum atraso no processamento cognitivo ou distúrbio de aprendizagem por volta dos nove anos de idade. Destaque é dado para a dislexia, uma vez que 85% dos disléxicos possuem ou já apresentaram comprometimento da linguagem oral na infância.

Invariavelmente, de acordo com a ASHA (2007), os problemas relativos à linguagem oral, especialmente as alterações de compreensão de palavras, frases e textos ou de integração dos significados de palavras em sentenças e destas em textos, estão presentes também quando há dificuldades posteriores de leitura e escrita. Diversas pesquisas têm relacionado o desenvolvimento da linguagem oral com o posterior desenvolvimento da leitura e escrita, salientando que as dificuldades de aprendizagem observadas em alunos a partir do ingresso no ensino formal, podem estar associadas ao comprometimento da linguagem oral manifestada antes desta fase de escolarização (Capovilla & Dias, 2008; Gómez, Duarte, Merchán, Aguirre, & Pineda, 2007, Prieto e cols., 2008).

Deste modo, problemas na alfabetização de crianças podem ser prevenidos com a avaliação da compreensão da linguagem oral logo no início da escolarização. Isto porque,

detectando-se precocemente transtornos de linguagem oral é possível implementar intervenções que minimizem a incidência ou o agravamento de dificuldades na leitura e na escrita, modificando sensivelmente o prognóstico (Capovilla, e cols., 2006, Pedroso & Rotta, 2006). Salienta-se que a avaliação e o diagnóstico em linguagem oral devam ser baseados em critérios que auxiliem os profissionais a caracterizarem o desempenho e identificar os aspectos deficitários do avaliando, tanto no que se refere à compreensão do que ouve, quanto na forma como ele se expressa. A identificação de áreas específicas que apresentam déficits deve auxiliar no direcionamento das condutas terapêuticas (Chevrie-Muller & Narbona, 2005). Com base nessas premissas, são destacadas a seguir as características que envolvem a avaliação da linguagem oral.

1.2 Avaliação da Linguagem

Avaliar a linguagem implica em investigar as facetas envolvidas no processo de comunicação, seja este processo realizado por meio de sons, gestos ou escrita. Há que se considerar, invariavelmente, a complexidade do sistema linguístico, que envolve estruturas cerebrais e sensório-motoras, funções cognitivas, bem como as influências de ordem social e escolar (Acosta, Moreno, Ramos, Quintana & Espino, 2003; Narbona & Fernández, 2005). Deste modo, Dockreel e McShane (1993), referindo-se à investigação das dificuldades de linguagem, reforçam que o sistema linguístico deve ser tratado como uma série de componentes que podem ser prejudicados em uma maior ou menor extensão mais do que como se fosse um sistema único. Acrescentam ainda, que, por esta razão, a criança deve ser o foco central da avaliação, que, por sua vez, deve ser formada tendo como base um modelo de sistema linguístico.

Considerando que a avaliação da linguagem oral realizada sob o foco psicolinguístico, aqui utilizado como referência, é dirigida à verificação do processo psicológico de compreensão e expressão da linguagem falada, com foco na competência linguística e comunicativa, tal avaliação implica na consideração de todos os processos envolvidos, desde o nível sensorio-motor (vias eferentes e aferentes) até os aspectos morfossintáticos, semânticos e mais amplamente cognitivos da linguagem (Chevrie-Muller, 2005).

Chevrie-Muller (2005) afirma que o estabelecimento de um modelo psicolinguístico, que leve em conta os aspectos psicolinguísticos e neurolinguísticos, permite que a avaliação e a identificação de mecanismos subjacentes à linguagem, quer seja em funcionamentos normais ou patológicos. O autor cita que diferentes pesquisadores, como Kirk e colaboradores (1968); Osgood (1957); Gérard (1991 e 1993); Aram e Nation (1982), entre outros, propuseram alguns modelos e instrumentos voltados a verificar a linguagem, contudo, ele considera que tais modelos não permitem uma descrição completa dos aspectos que envolvem a linguagem.

Nesse sentido, diante da necessidade de referência a um modelo que englobe tais aspectos, Chevrie-Muller e Narbona (2005) propuseram o *Modelo Neuropsicolinguístico* (MNPL) com a intenção de estabelecer um modelo capaz de avaliar o processamento da linguagem oral considerando os seguintes critérios: (a) as modalidades auditivo-orais, (b) a descrição da fala e da linguagem, (c) os aspectos anátomo-funcionais, (d) o estabelecimento da interface entre o comportamento linguístico observável, dados neurolinguísticos e dados psicolinguísticos, e (e) o auxílio na compreensão da patologia.

Assim, o esquema MNPL comporta uma entrada auditiva e uma saída oral (critério 'a'), permite a descrição dos diferentes níveis de funcionamento da linguagem (critério 'b'). Os aspectos formais (fonológico, morfossintático, lexical) e o nível semântico pragmático

nas duas vertentes da compreensão e da expressão. O aspecto anátomo-funcional e a interface entre o processo linguístico e o processo neurológico também são representados (critérios 'c' e 'd').

Como pode ser observado na Figura 1, o MNPL está organizado em dois eixos verticais (recepção/compreensão e expressão), três níveis de tratamento linguístico na horizontal (primário ou sensorio-motor, secundário ou de integração gnósica e terciário ou das operações cognitivas) e em duas colunas externas, à direita e esquerda, que representam as estruturas nervosas que fundamentam o processo linguístico nos diferentes níveis. Tendo em vista que o presente estudo tem como foco a compreensão da linguagem oral, o dar-se-á preferência para o detalhamento do processamento da recepção do sinal da fala, conforme descrito na sequência.

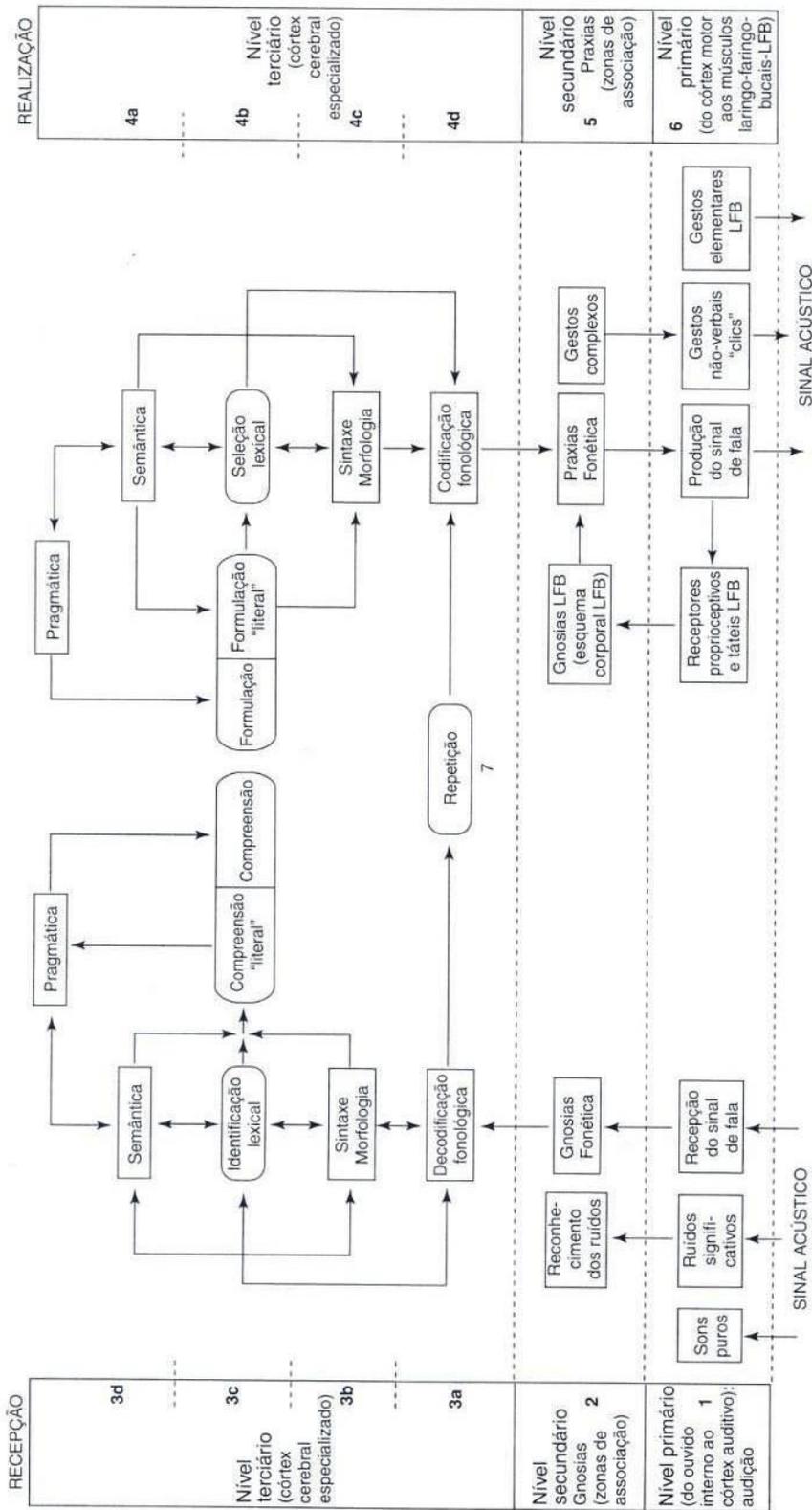


Figura 1. Modelo neuropsicolinguístico – MNPL (Chevrie-Muller & Narbona, 2005, p.104)

A compreensão do esquema MNPL referente à linguagem receptiva, se dá seguindo uma sequência ascendente, partindo-se da entrada do sinal acústico, como pode ser

observado da Figura 1, e seguindo as diversas unidades de tratamento da linguagem. No Nível Primário (1), ou sensorio motor, representado pelos órgãos sensoriais, audição, cinestesia, propriocepção, músculos e controle motor, ocorre o processamento da informação do ouvido interno até o córtex auditivo, ou seja, a captação de um sinal acústico, que pode ser um som puro, um ruído significativo ou um sinal da fala.

Ascendendo-se no processamento da informação, passa-se para o Nível Secundário (2), a unidade de tratamento de integração gnósica fonética e práxica, que ocorre nas zonas cerebrais de associações (envolvendo o córtex associativo secundário), momento em que ocorre o reconhecimento de um ruído ou da fala. Quando é reconhecido um sinal correspondente à fala a próxima etapa, já no Nível Terciário (3), que é o nível das operações cognitivas linguísticas ocorridas no córtex cerebral especializado, advém inicialmente a “decodificação fonológica” (3a), isto é, a identificação dos sons como segmentos que compõem as palavras ouvidas.

Posteriormente, a próxima unidade de tratamento se refere à “sintaxe” e à “morfologia” (3b), processo que se refere à combinação regras gramaticais que regulam a união de sílabas em palavras e destas em frases. Esta unidade pode levar à identificação lexical, que corresponde à significação da palavra que foi ouvida, a partir do acionamento do conhecimento prévio que o ouvinte possua. A “identificação lexical” (3c) pode levar à compreensão literal ou acessar a “Semântica” (3d). A compreensão literal permite entender o significado de cada palavra ouvida, contudo, uma sentença é mais do que um conjunto de palavras independentes agrupadas. Para que se dê sentido à sentença ouvida é preciso que seja acessada a unidade de tratamento denominada “Semântica” (3d), que engloba as regras gramaticais para a composição de frases, orações e sentenças. Tanto a semântica quanto a compreensão literal, quando aliadas à “Pragmática”, que se refere à contextualização do que foi ouvido, levam à efetiva compreensão da linguagem oral. Apesar de não ser

mencionada no MNPL a prosódia está presente no nível terciário, contemplando fatos como a entonação de voz, a acentuação, a velocidade da fala e seu ritmo.

Cabe salientar que, a partir do acesso ao nível Terciário, é possível observar na Figura 1 que as setas passam a ser bidirecionais. Segundo os autores Chevrie-Muller e Narbona (2005), isto ocorre porque não se pode conceber o processo de maneira puramente hierárquica e linear, pois no que tange a compreensão da linguagem (recepção) as unidades de tratamento da informação trabalham de modo encaixado e não independente. O autor exemplifica que a decodificação auditiva de uma palavra dotada de sentido, que faça parte do léxico do ouvinte, será mais rápida do que a de uma palavra desconhecida, pois o ouvinte pode reconhecê-la antes de ter identificado previamente cada fonema.

Diante do exposto, levando em conta que a linguagem oral está organizada nos aspectos fonológico, sintático, semântico e pragmático, e ponderando que ela é constituída por diferentes habilidades como nomeação, repetição, fluência, consciência fonológica e consciência sintática, entre outras, já existem instrumentos brasileiros destinados à avaliação destas habilidades, conforme levantamento realizado por Capovilla e cols. (2006). Entre estes instrumentos estão a Lista de Avaliação de Vocabulário Expressivo – LAVE, que avalia quais palavras uma criança fala; o Teste de Vocabulário por Imagens *Peabody*, que avalia o vocabulário receptivo; o Teste Infantil de Nomeação, que avalia a capacidade de nomear figuras; a Prova de consciência Fonológica, que avalia a capacidade de manipular segmentos da fala; e a Prova de Consciência Sintática, que está diretamente correlacionada ao desempenho em consciência fonológica, leitura, escrita e vocabulário.

Contudo, mais recentemente, Gurgel, Plentz, Joly e Reppold (2010) revisaram sistematicamente as bases de dados eletrônicas Medline, Lilacs, PsycInfo e Biblioteca Cochrane, com o objetivo de verificar os principais instrumentos utilizados em estudos de ensaio randomizado para avaliação da compreensão da linguagem oral em crianças no

período de janeiro de 1980 a março de 2010. Como termos de busca utilizaram: “*Language Tests*”, “*Comprehension*” e “*Child Language*” e seus entre termos.

A busca retornou 889 resumos, a partir dos quais somente 7 estudos foram incluídos, sendo que os demais foram descartados por inclusão de indivíduos que não estavam entre a faixa etária de zero a 18 anos e definição ausente ou incompleta dos autores do que foi considerado compreensão da linguagem. Os instrumentos utilizados nos artigos selecionados foram *Peabody Picture Vocabulary Test*, *Peabody Picture Vocabulary Test-Revised* (PPVT-R), *Swedish Communication Screening at 18months of age* (SCS18), *Test for Reception of Grammar - 2* (TROG-2), *Reynell Test*, *Reynell Development Language Scales* e *Reynell Developmental Language Scales-II*. Os pesquisadores constataram, pois, a escassez de instrumentos disponíveis para avaliação da linguagem oral, sendo que a maioria do que foi encontrado refere-se unicamente à avaliação do vocabulário receptivo. Além disso, constataram que são poucos os que possuem estudos de busca por evidência de validade.

Estes dados reforçam a necessidade de pesquisas que visem à construção de instrumentos destinados à avaliação da compreensão da linguagem, como é o caso do presente estudo, que explora a linguagem receptiva oral, que, conforme já mencionado, se refere aos canais de entrada da informação, que comportam a audição e a compreensão da linguagem falada (Gil, 2002). Sob a perspectiva da avaliação psicolinguística, foi construída a Bateria Informatizada de Linguagem Oral – BILO (Joly, 2008) que tem como foco avaliativo a compreensão da linguagem oral, estando pautada no exame do que a criança é capaz de entender a partir do que ouve. Esta bateria engloba, pois, a avaliação das várias unidades de processamento da informação descritas no MNPL, sobretudo os aspectos morfossintáticos, lexicais e semântico-pragmáticos.

Ressalva há que ser feita ao fato de que, no que concerne ao desenvolvimento e uso de instrumentos de testagem psicológica, pressupõe-se que sejam tomados cuidados metodológicos que assegurem que o teste possua qualidades psicométricas que lhe confirmem evidências de validade, de precisão e normatização (Anastasi & Urbina, 2000). Ao lado disso, em se tratando de testagem psicológica, há considerações específicas a serem descritas.

1.3 Características dos instrumentos de avaliação

A avaliação psicológica tem como objetivos obter informações a respeito de diferentes dimensões psicológicas do indivíduo ou grupo, e, por conseguinte, levar a uma maior compreensão dos mesmos. Trata-se de um processo de coleta de dados por meio de métodos e instrumentos psicológicos, tais como observações, entrevistas e testes, visando a formulação das interpretações de informações, que irão auxiliar o psicólogo nos processos de tomada de decisões e no planejamento de intervenções adequadas à cada pessoa (Weschler & Guzzo, 2005). Entre os instrumentos psicológicos que auxiliam a avaliação estão os testes informatizados, ou seja, aqueles nos quais a organização e a aplicação são realizadas por aplicativos instalados no computador ou por sistemas acessados via internet (Adanés, 1999; Joly & Noronha, 2006).

Considera-se que, na testagem psicológica estes testes incrementam as técnicas avaliativas, porquanto fornecem recursos visuais, sonoros e multimídias que permitem a simulação de situações que levam os testandos a apresentarem suas habilidades em condições mais realistas. Tais situações e condições mais autênticas seriam difíceis ou impraticáveis em testes do tipo lápis-papel padronizados (Chung & Baker, 1997 citado por Wall, 2000; Reynolds, Livingston & Wilson, 2009).

Além disso, os testes informatizados permitem a adaptação para pessoas com diferentes tipos de deficiência. Exemplos são os escritores de texto que podem ajudar pessoas com deficiência visual a responderem questões de modo oral; ou a tecnologia de reconhecimento de voz ou de detecção de movimentos leves da cabeça ou olhos que podem permitir que pessoas fisicamente comprometidas possam se expressar e serem avaliadas. Aquelas com dificuldades no controle motor fino podem usar o toque em tela ou lousa digital para responder os itens da avaliação (Wall, 2000).

Apesar dos elementos positivos descritos, é preciso considerar que existem potenciais problemas que acompanham o uso da tecnologia em avaliação. Entre estes, pessoas sem acesso às Tecnologias de Informação e Comunicação podem ser desfavorecidas com testes informatizados, pois aumentaria a dificuldade de execução da tarefa, ao passo que testandos acostumados com o uso de tecnologias poderiam ser favorecidos (Wall, 2000). Joly e Noronha (2006) acrescentam que a falta de habilidades em tais tecnologias também pode ser um elemento dificultador. Há também a questão sobre a falta de contato humano significativo, uma vez que, em testes informatizados há pouco ou nenhum contato com o avaliador, o que poderia dificultar a compreensão da tarefa, pois um orientador poderia ajudar um testando em situações de dúvidas, entaves do sistema ou na interpretação dos resultados (Wall, 2000).

Especialmente no que tange aos testes que são realizados pela Internet existe a preocupação sobre a identidade das pessoas avaliadas, tanto no que se refere à necessidade de certificação de que os respondentes são quem dizem ser; quanto pelo compromisso do avaliador em garantir que as informações do respondente serão mantidas em sigilo. Também, há a preocupação com a qualidade dos testes informatizados no que se refere às informações técnicas necessárias para mostrar o grau de qualidade e adequação do teste e às suas propriedades psicométricas. Se o teste já possui uma versão em lápis-papel validada, é

preciso garantia de que houve estudos de comparabilidade que lhe garantam validade (Wall, 2000).

Mapeando as publicações referentes à Avaliação Psicológica Informatizada no período de 2000 a 2004, Joly, Martins, Abreu, Souza e Cozza (2004), analisaram a produção científica brasileira e internacional disponibilizada na base de dados *Psycinfo* e na Biblioteca Virtual da Saúde (BVS-Psi). Foram encontradas 559 publicações sobre Avaliação Psicológica, das quais somente 17 artigos e 4 capítulos de livros referiam-se à Avaliação Psicológica Informatizada. Apesar de evidenciada a pouca utilização de recursos tecnológicos em testagem psicológica (1%), cabe destacar que, nesta área específica a produção brasileira apareceu como equivalente à internacional.

Mais recentemente, Joly e Reppold (2010) organizam um livro, em formato eletrônico, com pesquisas relacionadas ao desenvolvimento de instrumentos informatizados no Brasil. São apresentados no livro os seguintes instrumentos: a Bateria Informatizada de Linguagem Oral (BILO); o Teste de Wisconsin de Classificação de Cartas –WCST - versão computadorizada (funções executivas); o Teste de Stroop Computadorizado (atenção seletiva); o Teste de Geração Semântica (controle inibitório); o Teste de Fluência Verbal – FAS (fluência verbal); o Simon Task (controle inibitório e memória de trabalho); o Teste de Stroop para pré-escolares (controle inibitório), o Teste de Atenção Contínua (controle inibitório e atenção sustentada); o Iowa Gambling Task - IGT (processo de tomada de decisão); o Teste de Visualização Espacial TVZ2006A (habilidade viso-espacial), o Teste Dinâmico de Leitura – TDL (compreensão de leitura); o Teste Dinâmico Informatizado para Avaliar o Raciocínio Indutivo (TEDRI); o Teste de Compreensão das Emoções - TEC (Competência emocional). Também são apresentados o Teste Informatizado de Percepção de Emoções Primárias – PEP (inteligência emocional); a Bateria de Provas de Raciocínio - BPR-5 e a Bateria Fatorial de Personalidade - BFP computadorizadas; o Inventário de

Habilidades Sociais, o Inventário de Habilidades Sociais para Adolescentes e o Inventário Multimídia de Habilidades Sociais para Crianças (habilidades sociais), além da Escala de Atitudes – eSATS.

Embora o uso de testes psicológicos informatizados ainda seja restrito, conforme observado, a área de construção destes testes tende a avançar rapidamente, tornando esta modalidade de testagem um recurso científico de qualidade, considerando as especificidades de validade, precisão e fidedignidade (Joly & Noronha, 2006). Isto porque, considerando que um teste psicológico deve fornecer uma medida padronizada (Anastasi & Urbina, 2000), os testes informatizados são recursos que proporcionam maior uniformidade no procedimento de aplicação e pontuação. No tocante à avaliação da compreensão da linguagem oral, a testagem informatizada permite que as instruções orais sejam padronizadas e minimizem-se os ruídos ambientais - que podem configurar variáveis intervenientes – com a utilização de fones de ouvidos.

Porém, para que os testes psicológicos sejam reconhecidos como instrumentos científicos e de segurança quanto a sua utilização, devem ser seguidos alguns critérios específicos na sua construção, como definição dos objetivos, especificação do contexto, construção dos itens e das instruções, revisão da primeira versão do teste por peritos, estudo-piloto, seleção de amostras e administração do teste inicial, análise e seleção empírica dos itens, avaliação da confiabilidade e da validade do teste, elaboração de normas e redação do manual para uso (Adanéz, 1999; Joly & Noronha, 2006). No tocante a testes informatizados, Joly e cols (2005) desenvolveram o Sistema de Avaliação para Testes Informatizados (SAPI), cujo objetivo é analisar e caracterizar instrumentos de avaliação psicológica informatizada.

De acordo com o SAPI, os aspectos que devem estar presentes em um instrumento de avaliação psicológica informatizada são os que dizem respeito às características

psicométricas, às questões técnicas e aos níveis de controle. Os aspectos relacionados às características psicométricas, referem-se aos itens específicos da avaliação psicológica, isto é, os relativos à construção do teste e suas qualidades psicométricas. Os aspectos referentes às questões técnicas fornecem as especificações sobre o equipamento (*hardware* e *software*); sobre a capacitação técnica para realizar a avaliação e para a utilização do equipamento; e sobre a utilização por pessoas com deficiência. Consideram-se aspectos que aludem aos níveis de controle, as condições de aplicação, necessidade de supervisão, treino e demonstração antecipada dos itens que compõem a avaliação, autenticidade da identidade do avaliando e possíveis fraudes, além da transferência de dados (Joly e cols, 2005).

Mais especificamente em relação aos aspectos ligados às características psicométricas, cabe destaque aos procedimentos de busca por validade e precisão do instrumento e ao estabelecimento de normas para a interpretação dos resultados. O estudo de validade de um teste constitui-se uma etapa fundamental, uma vez que permite atestar o grau de representatividade do domínio teórico, ou seja, fornece evidências de que os escores obtidos por meio do teste fazem referência aos pressupostos teóricos em que o instrumento se baseia, possibilitando a interpretação correta do fenômeno observado. Por isso, para se estabelecer a validade de um teste, deve-se considerar o uso específico para o qual o teste está sendo considerado e a conceituação teórica subjacente (Anastasi & Urbina, 2000). Segundo a *American Educational Research Association - AERA*, *American Psychological Association - APA* e a *National Council on Measurement in Education - NCME* (1999) as fontes de evidência de validade podem ser baseadas 1) no conteúdo de teste; 2) no processo de resposta; 3) na estrutura interna; 4) na relação com outras variáveis e/ou 5) na consequência de testagem, conforme serão descritas a seguir.

A evidência de validade com base no conteúdo de teste visa verificar a representatividade dos itens do teste, ou seja, se eles são amostras do construto que se

pretende avaliar com o teste. A obtenção desta evidência pode ser realizada por meio do julgamento de especialistas, que avaliam se há relação do teste com o construto a ser investigado. A evidência com base no processo de resposta tem por finalidade analisar os processos mentais envolvidos na execução do teste. Para tanto, pode ser feita a análise individual das respostas, solicitando aos testandos que expliquem as estratégias que utilizaram para responder ao teste; ou ainda pela análise da relação entre diferentes partes do teste ou o teste com outras variáveis. Essa validade também pode incluir estudos empíricos de como observadores ou juízes que registram e interpretam os dados a fim de verificarem se são coerentes com o que está sendo medido.

A evidência com base na estrutura interna do teste verifica a coesão dos itens do teste, levando em consideração as correlações entre itens que avaliam o mesmo construto e também as correlações entre subtestes avaliando construtos similares. É possível verificar, deste modo, se há coerência dos itens/subtestes com a estrutura proposta pela definição teórica. Esta estrutura conceitual pode sugerir uma dimensão ou mais dimensões (multidimensionalidade), também pode pressupor uma estrutura hierárquica ou de fatores isolados.

A evidência com base na relação com outras variáveis consiste na análise da relação dos escores do teste com variáveis externas. As variáveis externas são consideradas como medidas de critério do que é, teoricamente, esperado estar relacionado ao construto medido pelo teste, ou ainda podem ser medidas obtidas por outros testes que mensurem o mesmo construto, construtos relacionados ou construtos diferentes. Esta evidência de validade pode ser obtida por dois procedimentos: convergente-discriminante ou relação teste-critério. A evidência convergente-discriminante indica relação entre escores do teste e medidas de construtos similares (convergentes) ou diferentes (divergentes). Já a evidência baseada na relação teste-critério indica se os escores do teste são capazes de prever determinado

critério como, por exemplo, o critério desenvolvimental verificado com a idade e a escolaridade.

Por fim, a evidência baseada na consequência de testagem consiste em verificar se a utilização do teste está surtindo os efeitos desejados de acordo com o propósito para o qual foi criado. Deste modo, ocorre pelo exame das consequências sociais do uso do teste.

Em relação ao processo de fidedignidade, este se refere ao quanto um escore obtido na avaliação se aproxima das características estáveis do sujeito, ou seja, características constantes em seus comportamentos em situações cotidianas. Desta forma, os estudos de fidedignidade possibilitam aferir o tamanho do erro que geralmente ocorre nas avaliações, fornecendo maior confiabilidade para os resultados do teste. A fidedignidade pode ser definida pelo coeficiente de correlação (r) que expressa o grau de relacionamento entre dois conjuntos de escores. Este coeficiente deve se aproximar de '1' para que um teste seja confiável, sendo que a variabilidade possível está compreendida entre 0 e 1 (Anastasi & Urbina, 2000).

Já no que tange ao estabelecimento de normas para a interpretação dos resultados, consideram-se os estudos inerentes à padronização e à normatização. A padronização alude à uniformidade de procedimento na aplicação, correção e interpretação do teste, enquanto que a normatização refere-se ao estabelecimento de normas construídas empiricamente, dispondo sobre o desempenho normal ou médio da amostra de padronização em um determinado instrumento (Anastasi & Urbina, 2000).

Dito isto, considerando-se os aspectos envolvidos na avaliação psicológica informatizada, o presente estudo buscará evidências de validade para a Bateria Informatizada de Linguagem Oral (BILOv3) destinada a avaliar a compreensão da linguagem oral em crianças cursando o Ensino Infantil e Fundamental. Cabe realçar que, no concernente às avaliações de linguagem, há grande contribuição da avaliação

informatizada, uma vez que, pode-se usar áudio e vídeo para simular várias situações, garantindo padronização da pronúncia das palavras, uniformidade do som e entonação de voz das consignas. Com o uso de fones de ouvidos, a testagem pode ser realizada de modo coletivo, de acordo com o número de computadores disponíveis, cada testando em seu equipamento respondendo em seu próprio ritmo, com a minimização das influências de ruídos externos. Além disso a tecnologia permite que crianças que ainda não sabem escrever respondam às questões usando o mouse ou o toque em tela.

Adicione-se a estas vantagens a capacidade de registrar a rapidez do avaliando em reagir ou tomar decisões, posto que permite quantificar precisamente a latência média de resposta, ou seja, o tempo de intervalo entre o aparecimento do estímulo e a resposta. Há que se adverter que, em se tratando da avaliação de crianças, o computador não exclui a necessidade do contato humano do avaliador, pois este deve estar presente para assegurar que as crianças estejam ouvindo em um volume adequado, tenham compreendido a tarefa, saibam manusear o mouse e estejam aplicadas na realização do solicitado.

A BILO é um instrumento que se encontra na terceira versão, sendo que os estudos anteriores para a versão 1 e para versão 2 registram os estudos de validação no processo de construção e desenvolvimento, pelos quais o instrumento passou até a consolidação da versão atual (BILOv3). A seguir constam os estudos de evidências de validade e precisão aos quais a BILO foi submetida.

1.4 Estudos com a Bateria Informatizada de Linguagem Oral - BILO

A BILOv1 foi composta por sete subtestes/provas que objetivavam avaliar distintas habilidades requeridas para a compreensão da linguagem receptiva, são eles: Compreensão Morfossintática “CM” – 10 itens; Sequencia Lógica “SL” - 10 itens, Organização Lógico

Verbal “OLV” – 10 itens; Interpretação de Histórias “IH” – 9 itens; Completar Frases “CF” – 10 itens; Completar Parágrafos “CP” – 9 itens e Completar Histórias “CH” – 13 itens. A partir da segunda versão, foi retirada a prova CP. Salienta-se que, com a mudança do Ensino Fundamental de oito para nove anos entre 2008 e 2009, houve a alteração da nomenclatura “série” para “ano” de escolaridade. Optou-se, na descrição dos estudos, por manter a nomenclatura que cada autor utilizou na época da realização de seu estudo de modo a assegurar uma descrição mais autêntica. Portanto, quando for utilizado o termo “série”, remete-se ao período de oito anos de ensino fundamental e quando for utilizado o termo “ano” ao período de 9 anos.

Na versão 1, Joly e cols. (2008) visaram buscar evidências de validade de conteúdo e de construto para a Bateria Informatizada de Linguagem Oral – BILO, realizaram um estudo que contou com duas fases. A primeira fase do estudo tratou-se de uma análise de conteúdo realizada por juízes para buscar evidência de validade baseadas no conteúdo, e a segunda fase constou de uma análise fatorial com informação total por prova com objetivo de evidenciar validade com base na estrutura interna.

Participaram da primeira fase do estudo dois especialistas em Psicologia, dois especialistas em Tecnologias da Informação e Comunicação (TIC), dois professores do Ensino Fundamental e 10 estudantes por nível de ensino ($N=30$). Foram utilizados como instrumentos de avaliação o Sistema de Avaliação Psicológica Informatizada (SAPI) de Joly e cols. (2005), e entrevistas semi-estruturada com discentes e docentes. Como procedimento, houve a aplicação assistida da BILO em 30 crianças (10 alunos de cada série - pré, 1ª e 2ª série) e avaliação do instrumento por juízes. Os resultados do estudo de validade de conteúdo revelaram concordância dos avaliadores – especialistas em TIC - quanto a BILO ser um teste informatizado. Também, por meio da análise do conteúdo específico para a construção do instrumento, realizada pelos especialistas de Psicologia, foi

confirmado que a BILO afere habilidades das dimensões da linguagem oral (sintaxe, morfologia, semântica e compreensão oral). Não foram encontradas dificuldades para compreensão das instruções, reconhecimento das imagens e vocabulário pelas crianças; e além disso os professores concordaram que as atividades estavam apropriadas para as crianças. A única dificuldade encontrada foi em relação ao uso do computador, a qual foi ajustada com a inserção no programa de um tutorial-treino em habilidades básicas de informática.

Na segunda fase do estudo, Joly e cols. (2008), buscando a validade de construto, realizaram a análise fatorial com informação plena, considerando a Teoria de Resposta ao Item – TRI para a análise dos itens. As análises foram realizadas sobre os escores obtidos quando da aplicação coletiva da BILO em 125 crianças, com faixa etária variando de cinco a 10 anos ($M = 6,90$; $DP = 1,31$), que cursavam o último ano da educação infantil (pré) e a 1ª e 2ª série do Ensino Fundamental de escolas paulistas da Rede Pública de Ensino. Tendo em vista que cada prova avalia uma habilidade específica da Linguagem oral, a análise fatorial foi realizada por prova. Os resultados indicaram a unidimensionalidade dos itens de cada prova, no entanto houve a necessidade da retirada de alguns itens nas provas IH, CP e CH. A análise das qualidades psicométricas da BILO pela Teoria de Resposta ao Item, envolvendo todos os itens de cada prova, indicou ajuste ao modelo de três parâmetros com um bom índice de fidedignidade, variando de $KR=0,64$ à $KR=0,97$. Os itens de cada subteste foram classificados em função da dificuldade que apresentaram e da habilidade necessária dos participantes para responderem, considerando-se as respostas ao acaso o que sugeriu que alguns itens fossem revistos. Verificou-se correlação positiva e moderada entre todos os subtestes e destes em relação à BILO ($r_{CM} = 0,50$; $r_{SL} = 0,92$; $r_{OLV} = 0,85$; $r_{IH} = 0,62$; $r_{CF} = 0,68$; $r_{CP} = 0,58$; $r_{CH} = 0,76$, $p = 0,01$).

Em sequencia, visando ampliar as características psicométricas da BILOv1, Almeida e Joly (2008) buscaram evidências de validade para a BILOv1 correlacionando-a com o desempenho em vocabulário, verificado por meio do Teste de Vocabulário por Imagens *Peabody* (TVIP). Participaram do estudo 106 alunos de ambos os sexos, com idade entre cinco e oito anos, matriculados no último ano da educação infantil e primeiro ciclo do Ensino Fundamental (1ª e 2ª séries). A aplicação ocorreu em duas sessões de modo coletivo. Os resultados indicaram correlação moderada, positiva e estatisticamente significativa tanto entre cada prova da BILO quanto do seu escore total ($r = 0,431$, $p \leq 0,001$) com o desempenho dos participantes no TVIP. A partir das correlações as pesquisadoras apontaram a evidência de validade convergente para a BILOv1. Além disso, houve discriminação do desempenho na BILO das maiores e menores pontuações em função da amplitude do vocabulário (verificada pelos escores do Peabody), garantindo a evidência de validade de critério por grupos extremos.

No mesmo sentido, Issa (2008) objetivou encontrar evidências de validade para a BILOv1, correlacionando-a com a Escala de Transtornos de Déficit de Atenção e Hiperatividade para Professores. Participaram do estudo 125 alunos, com idades entre 5 e 8 anos, de escolas públicas que freqüentavam o último ano da educação infantil e a 1ª e a 2ª séries do Ensino Fundamental. A BILOv1 foi aplicada coletivamente nos alunos em sala de informática e os dados referentes ao déficit de atenção foram baseados nas respostas dos professores. O estudo revelou evidência de validade por critério de escolaridade ($F[2, 85]=7,77$; $p=0,001$) e por critério por mudanças desenvolvimentais ($F[5, 82]= 3,91$; $p \leq 0,05$) para a BILOv1. A análise correlacional indicou que houve correlação fraca negativa entre os subtestes da BILO Sequencia Lógica e Organização Lógico Verbal e os fatores déficit da atenção ($r_{SL} = -0,20$, $r_{OLV} = -0,23$; $r_{total} = -0,19$) e dificuldades de aprendizagem

($r_{SL} = -0,26$, $r_{OLV} = -0,27$; $r_{total} = -0,22$), evidenciando, segundo a pesquisadora, evidência de validade por critério divergente para a BILOvI.

Com a finalidade de buscar relação entre a compreensão da linguagem oral e a capacidade de raciocínio geral, Joly e Piovesan (2012), avaliaram 82 alunos que cursavam pré (31,7%), 1ª série (32,9%), 2ª série (35,4%), com idades entre 5 e 9 anos, sendo 52,4% do gênero masculino. Aplicaram a BILOvI, de forma coletiva em uma sessão, e em outra a Escala de Maturidade Mental Columbia, individualmente. Os resultados da ANOVA apontaram diferença estatisticamente significativa para os subtestes da BILO Compreensão Morfossintática ($F[2, 79] = 4,87$; $p = 0,01$), Sequencia Lógica ($F[2, 79] = 6,27$; $p = 0,00$) e Organização Lógico-verbal ($F[2, 79] = 3,16$; $p = 0,04$), em relação à série, diferenciando escores menores para as séries iniciais. A análise correlacional entre os escores dos dois instrumentos revelou correlações baixas e estatisticamente significativas ($r_{SL} = 0,28$; $r_{OLV} = 0,30$, $r_{IH} = 0,30$; $r_{total} = 0,29$) entre a compreensão da linguagem oral e a capacidade de raciocínio geral. Foi realizada a análise de regressão múltipla, que revelou o raciocínio geral como representante de 29% das habilidades necessárias para um bom desempenho na BILO. As pesquisadoras consideraram os resultados como mais uma evidência de validade para a bateria.

Joly e Dias (2009) visando buscar evidências de validade de construto para a BILOvI, realizaram uma análise fatorial de cada prova com informação plena nos escores de uma amostra de 119 crianças. Os respondentes tinham de cinco a dez anos, e cursavam o pré, a 1ª e 2ª série de escolas paulistas da Rede Pública de Ensino. A BILOvI foi aplicada nos laboratórios de informática das escolas. Os resultados obtidos por meio da análise fatorial revelaram que os itens de cada subteste indicaram boas cargas fatoriais para as provas $CF > 0,60$; $SL > 0,50$; $OLV > 0,70$; $CM > 0,30$, contudo nas provas IH, CH e CP foi necessária a retirada de alguns itens durante a análise para que as provas ficassem com as

seguintes cargas: $IH > 0,30$; $CH > 0,40$; $CP > 0,50$. Apesar de comprovada a unidimensionalidade de cada prova, denotando que são aferidas habilidades específicas de linguagem em cada uma das provas, constatou-se a necessidade de se rever a estrutura da prova IH, com a alteração das imagens dos itens 3, 5, 6 e 8; e das provas CP e CH, que deveriam ser reduzidas a uma única prova. Sugeriu-se, assim, que a prova Completar Histórias contemplasse os itens CP7, CP8 e CP9, CH1, CH2, CH3, CH4, CH5, CH6, CH7, CH8, CH9, CH10 E CH11 e que fossem retirados cinco itens da prova CF que se apresentaram como muito fáceis (itens CF1, CF2, CF5, CF7 e CF10) para a 2ª versão da BILO.

Considerando os estudos apresentados e as sugestões procedentes dos estudos de Joly e cols. (2008) e Joly e Dias (2009), a BILO foi alterada, sendo retirados itens da prova CF e excluída a prova Completar Parágrafos, cujos itens foram eliminados ou distribuídos na prova CH (Completar Histórias). Além disso, foi realizada a troca de algumas imagens e a alteração de instruções que dificultavam a compreensão da consigna. A BILOv2 passou a ser constituída por seis provas, quais sejam, Compreensão Morfosintática – CM (10 itens); Sequencia Lógica – SL (10 itens); Organização lógico-verbal – OLV (10 itens); Interpretação de Histórias – IH (9 itens), Completar Frases (5 itens); e Completar Histórias – CH (16 itens). Nota-se nesta segunda versão que a prova CF passou de 10 para 5 itens e que a CH passou de 13 para 16 itens. Passou-se às investigações para a busca de validade para a BILOv2.

Em relação à versão 2 da BILO, Joly e cols. (2009) buscaram evidências de validade correlacionando-a com o Teste de Vocabulário por Imagens Peabody (TVIP), assim como realizado com a BILOv1 (Joly & Almeida, 2008). Participaram do estudo 158 alunos de 1º a 3º anos do Ensino Fundamental, com idades entre 6 e 10 anos, sendo que, 51,3% eram meninos. Os participantes responderam individualmente à BILOv2 e coletivamente ao

Teste de Vocabulário por Imagens (TVIP). A MANOVA indicou diferença significativa para a idade e série em relação ao escore total da BILOv2, revelando influência do desenvolvimento e da aprendizagem na compreensão da linguagem oral. O escore total do TVIP correlacionou-se positiva e significativamente com os escores por provas da BILOv2 (TVIP e CM, $r=0,31$; $p=0,000$; TVIP e IH, $r=0,37$; $p=0,000$; TVIP e SL, $r=0,33$; $p=0,000$; TVIP e CF, $r=0,31$; $p=0,000$; TVIP e CH, $r=0,22$; $p=0,003$; TVIP e OLV, $r=0,50$; $p=0,000$, e com o total da BILOv2 ($r=0,42$; $p=0,000$), conferindo à BILOv2 evidência de validade convergente, corroborando os estudos de Joly e Almeida (2008). Além disso, o índice de fidedignidade da BILOv2 apontado pelo Alfa de Cronbach foi igual a 0,89, o que atribui características psicométricas adequadas à BILOv2.

Também buscando validade para a BILOv2, Joly, Piovezan, Soares, Lopes e Martins (2009) a fizeram por meio da correlação com o Teste de Competência de Leitura Silenciosa (TeCoLeSi). Para tanto, participaram do estudo 162 crianças de uma escola pública, sendo 51,2% do gênero masculino, que frequentavam do primeiro ao terceiro ano do Ensino Fundamental. A idade variou entre 6 a 10 anos ($M = 7,6$; $DP = 0,98$). O TeCoLeSi foi aplicado coletivamente na própria sala de aula de cada turma e a BILOv2 foi aplicada na sala de informática disponibilizada pela escola. Os resultados indicaram evidência de validade para a BILOv2, uma vez que, o escore total do TeCoLeSi correlacionou-se positiva e significativamente com os escores por provas e o total da BILOv2 ($r_{\text{Total}} = 0,442$; $p \leq 0,01$). Também obtiveram evidência de validade desenvolvimental para o nível de escolaridade ($F [2, 160]=9,245$; $p = 0,008$). Além disso, verificaram índice de fidedignidade igual a $\alpha = 0,92$.

Joly, Reppold e Dias (2009), visando comparar o desempenho em compreensão oral de crianças paulistas e gaúchas e verificar a precisão da bateria, aplicaram a BILOv2 coletivamente em 213 alunos com idade entre cinco e oito anos que cursavam, em escolas

públicas dos estados de São Paulo ($n = 81$; idade média igual a 6,91 anos, $DP=0,73$) e Rio Grande do Sul ($n = 132$, idade média igual a 6,83 anos, $DP=0,83$), o último nível de ensino da educação infantil, e primeiro ciclo do Ensino Fundamental (1ª e 2ª séries). A amostra estava distribuída equitativamente quanto ao sexo. A análise da diferenças de média entre as regiões pelo teste t de *Student* não revelou diferença estatisticamente significativa para os estados - CM ($t[211] = -1,230$; $p = 0,220$), IH ($t[211] = -1,916$; $p = 0,057$), CF ($t[211] = 1,259$; $p = 0,209$), CH ($t[211] = -0,416$; $p = 0,678$), SL ($t[211] = 1,189$; $p = 0,236$) e OLV ($t[211] = 1,246$; $p = 0,214$), o que sugeriu possibilidade de elaboração de normas comuns para os dois estados. Também foram verificadas diferenças significativas na amostra total por gênero, nas provas SL e OLV, nas quais as meninas apresentaram um desempenho superior aos meninos. Em relação à idade observou-se diferenças significativas entre as idades de 5 a 8 anos e entre as séries para todas as provas, demonstrando a influência do desenvolvimento e da série freqüentada na compreensão da linguagem oral, sendo que as médias aumentaram com o avanço da idade e da escolaridade.

Neste mesmo estudo de Joly, Reppold e Dias (2009), quando realizada a análise correlacional entre as provas da BILOv2 para as duas regiões separadamente, constataram associações significativas e com intensidade de baixa a alta nos resultados dos dois estados (SP e RS), sendo a magnitude das correlações por estado próxima, para a maioria das provas, exceção feita para CF e CH, sugerindo revisão técnica para controlar o padrão de respostas aos itens. Por meio do cálculo do Alfa de Cronbach, as pesquisadoras encontram bons índices de precisão para as provas CM (SP 0,69; RS 0,79); IH (SP 0,73; RS 0,68), CH (SP 0,73; RS 0,82); SL (SP 0,87; RS 0,88) e OLV (SP 0,92; RS 0,89); contudo, constataram que a prova CF (SP 0,12; RS 0,51) foi a única que revelou disparidade de magnitude do índice de precisão para as duas amostras e um índice considerado muito baixo para a precisão de uma prova (SP 0,12).

Tendo em vista os três estudos relativos a BILOv2, fez-se necessária a alteração das provas Interpretação de Histórias, Completar Frases e Completar Histórias, no que se refere ao número de itens que se mostravam díspares prejudicando as análises – a prova IH estava com 9 itens, a CF com 5 itens e a CH com 16 itens. Deste modo, foram redistribuídos os itens entre as provas CF e CH, e criado outro item para a prova IH de modo que cada prova ficou com 10 itens. Além disso, uma das imagens da prova CM, referente a uma logomarca, foi substituída por outra tendo em vista que tal logomarca não seria reconhecida em todos os estados brasileiros; também o item que continha a brincadeira “pular amarelinha” foi substituída pela brincadeira “pular corda”. Ainda, houve a reprogramação do sistema para que os participantes fossem impedidos de avançarem nas respostas das provas CF e CH sem que todas as lacunas de uma tela estivessem preenchidas. Foi acrescida ao programa a possibilidade de seleção de provas a serem realizadas. Outra adição à bateria foi a possibilidade de aferir o tempo gasto pelos respondentes por item, prova e total da BILO. Em virtude de tais modificações, a terceira versão da BILO foi gerada (July, 2009) e iniciou-se a busca por evidências de validade.

O primeiro estudo por busca de evidências de validade para a BILOv3 foi o de Istome (2010), que correlacionou os escores do instrumento ao desempenho em vocabulário, avaliado pelo Teste de Vocabulário por Imagens Peabody (TVIP) computadorizado. Participaram 157 alunos de ambos os sexos com idade entre quatro e sete anos, regularmente matriculados no último nível de ensino da educação infantil e primeiro ano do Ensino Fundamental de três escolas privadas. A aplicação ocorreu em duas sessões, a BILOv3 foi aplicada no primeiro dia e o TVIP uma semana após a primeira aplicação. Ambos os instrumentos foram aplicados coletivamente no laboratório de informática. As análises estatísticas indicaram que as médias de acertos aumentaram no 1º ano do ensino fundamental quando comparados com os escores do Ensino Infantil em quase todas as

provas e no escore total da BILO ($t[155] = -4,652; p \leq 0,000$), exceção para a prova Compreensão Morfosintática, que teve diferença marginalmente significativa ($t[155] = -1,866; p \leq 0,064$).

Não foram encontradas diferenças em relação ao gênero dos participantes para os escores da BILO e TVIP, contudo observou-se diferença significativa entre o tempo dos meninos em relação ao das meninas na BILOv3, sendo que elas utilizaram mais tempo na tarefa ($t[155] = 2,948; p = 0,004$). No tocante às diferenças por idade a análise de variância das médias foi obtida por meio da comparação de dois grupos etários: alunos com cinco anos ou menos e alunos com seis anos ou mais, por meio do Teste *t* de *Student*. Somente a prova CM mostrou-se marginalmente significativa ($t[152] = -1,885; p \leq 0,061$), sendo que, tanto as demais provas quanto o escore total da BILO ($t[153] = -4,257; p \leq 0,000$) tiveram diferenças significativas ($p \leq 0,05$) ou muito significativas ($p \leq 0,001$). O Teste de correlação de Pearson indicou correlação entre todos os escores por subteste da BILOv3 com o TVIP (TVIP e CM, $r=0,29; p=0,000$; TVIP e SL, $r=0,33; p=0,000$; TVIP e OLV, $r=0,39; p=0,000$; TVIP e IH, $r=0,38; p=0,000$; TVIP e CF, $r=0,27; p=0,000$; TVIP e CH, $r=0,37; p=0,000$), e com o total da BILOv2 ($r=0,48; p=0,000$), fornecendo evidência de validade baseada na relação com outras variáveis à BILOv3.

Também, Joly, Cruz, Lima, Prates e Leme (2010) buscaram evidências de validade da Bateria Informatizada de Linguagem Oral (BILOv3) por meio de critério desenvolvimental. Para tanto investigaram as diferenças de desempenho dos participantes em função da idade e do nível de ensino. Também foram verificadas diferenças por gênero e por tipo de escola frequentada- privada ou pública. A BILOv3 foi aplicada a 353 alunos de duas escolas (pública, $N= 199$ e privada $N = 154$), com idades entre 6 e 10 anos, que frequentavam o 2º ($N = 115$), 3º ($N = 80$), 4º ($N = 104$) e 5º ($N = 54$) anos escolares. As análises revelaram que houve diferença estatisticamente significativa no desempenho em

compreensão da linguagem oral em função da escolaridade para os alunos das escolas públicas em todas as provas ($F_{SL}[3, 150] = 16,89; p < 0,001$), enquanto que para os alunos das escolas particulares a diferença se deu somente nas provas IH e CF, diferenciando os alunos do 2º ano, com menor desempenho, dos alunos do 5º anos com melhor desempenho. Com relação à diferença de médias em função da idade, constatou-se que somente foi significativa para alunos da escola pública, indicando que os mais novos tiveram desempenho inferior ao dos mais velhos em todas as provas ($F_{SL}[4, 149] = 13,29; p < 0,001$), com exceção da CH.

Os autores ainda constataram que houve diferença significativa em função o tipo de escola freqüentada, sendo que o desempenho de estudantes de escola privada foi melhor do que o desempenho dos alunos da escola pública em todas as provas da BILOv3, com exceção da prova CF. Diferença significativa de médias em função do gênero só foi encontrada nas provas CM ($t[83,57] = 4,17; p < 0,001$) e SL ($t[119,85] = 2,37; p < 0,001$) nos escores dos alunos da escola particular, indicando que as meninas tiveram melhores médias nestas provas. Os autores concluíram que a BILOv3 foi sensível para discriminar estudantes mais velhos dos mais novos e de séries mais avançadas dos iniciantes.

Outro estudo realizado com a BILOv3, foi o de Freitas (2011) que buscou verificar evidências de validade por meio da correlação com o Teste Token computadorizado, que avalia a compreensão da linguagem por comando verbal. Participaram 101 estudantes do 1º ao 3º ano do ensino fundamental de uma escola particular de Natal, Rio Grande do Norte. As idades dos participantes variaram de 5 a 8 anos ($M=6,81; DP=0,89$) e 50,49% da amostra total era do sexo feminino. Os instrumentos foram aplicados coletivamente, em única sessão, no laboratório de informática da escola. As análises relativas à idade foram realizadas considerando duas faixas etárias (5-6 anos e 7-8 anos). Foram verificadas diferenças para o escore das provas SL, OLV, CF e também para escore total da BILO

($t[99] = 2,71; p \leq 0,01$) em relação à idade, sendo que os mais novos obtiveram menores pontuações em todos os casos. Em relação ao gênero dos participantes observou-se que as meninas obtiveram pontuações mais altas que os meninos, esta diferença foi estatisticamente significativa em quase todas as provas da BILOv3 e no escore total ($t[99] = 2,71; p \leq 0,01$), exceção feita às provas CM e IH em que as diferenças podem ser consideradas ao acaso. Não foram encontradas diferenças estatisticamente significativas para o tempo de execução, nem em função do gênero, nem do grupo etário.

A ANOVA revelou que o desempenho dos participantes nas provas CM, SL e OLV e no total da BILO diferenciaram significativamente do 3º ano em relação ao 1º e 2º. Com relação ao tempo, constatou-se que alunos do 1º ano usaram mais tempo do que os do 2º e 3º anos para a execução das provas. Em relação a correlação, foi verificada associação significativa entre as provas OLV ($r = 0,31; p = 0,002$) e SL ($r = 0,23; p = 0,022$) da BILO e o escore total do TOKEN. Foi verificada associação entre cada prova da BILO e escore total por meio da correlação de Pearson, que revelou associações altamente significativas. As provas CH, CF, OLV, SL apresentaram correlação acima de 0,80, enquanto que CM e IH apresentaram associações iguais a 0,70 e 0,79 respectivamente. Os índices de confiabilidade das provas da BILO foram calculados pelo *Alfa de Cronbach* sendo, CM = 0,68; SL = 0,87; OLV = 0,89; IH = 0,83; CF = 0,65 e CH = 0,43.

Soares (2011), buscou evidências de validade para a BILOv3 por meio da correlação com o desempenho escolar e com a competência em leitura silenciosa. Participaram 290 alunos de uma escola pública do estado de Macapá, matriculados no 1º, 2º, 3º, e 4º anos do Ensino Fundamental, sendo 54,1% do gênero masculino, com idade média igual 7,12 ($DP=1,23$ anos). Foram aplicados o Teste de Desempenho Escolar (TDE); o Teste de Competência de Leitura de Palavras e Pseudopalavras (TCLPP) e a Bateria Informatizada

de Linguagem Oral (BILOv3), em três sessões, com intervalo de aproximadamente uma semana.

Os resultados desta pesquisa revelaram que os melhores desempenhos foram para as provas Compreensão de Histórias e Compreensão Morfossintática, enquanto que os desempenhos mais baixos foram para as provas Sequência Lógica e Orientação Lógico-Verbal. Foram encontradas diferenças significativas de desempenho para todas as provas da BILOv3 e TCLPP em função da idade ($F_{\text{Total}} [4, 285] = 39,284; p < 0,001$) e do ano escolar ($F_{\text{Total}} [3, 286] = 59,061; p < 0,001$). As diferenças decorrentes do sexo do aluno não foram estatisticamente significativas ($t[288] = 1,918; p = 0,359$), indicando que podem ser atribuídas ao acaso. Foram encontradas associações fracas e positivas entre as provas da BILOv3 com as categorias do TCLPP “Palavra com incorreção semântica”, “Pseudopalavras com trocas visuais”, “Pseudopalavras com trocas fonológicas”, “Pseudopalavras estranhas” e com o total do TCLPP (CM, $r=0,225; p < 0,001$; SL, $r = 0,399; p < 0,001$; OLV, $r=0,371; p < 0,001$; IH, $r=0,202; p=0,001$; CF, $r=0,252; p < 0,001$; CH, $r = 0,195; p = 0,001$; Total, $r = 0,405; p < 0,001$). Assim, o pesquisador constatou evidências de validade para a BILOv3 em estudantes do Macapá.

Por fim, Cruz (2011) realizou um estudo psicométrico da BILOv3, buscando evidências de validade por meio da correlação com o Teste de vocabulário por Imagens *Peabody* informatizado (TVIP). Os dois instrumentos foram aplicados, coletivamente, em única sessão, no laboratório de informática de uma escola pública do interior do Estado de São Paulo. Responderam aos testes 325 crianças do 1º ao 5º ano, com idades entre 6 e 13 anos ($M = 9,03$ e $DP = 1,535$), sendo 49,5% dos participantes do sexo feminino.

Os resultados da análise correlacional do estudo de Cruz (2011) indicaram associações positivas entre as provas e total da BILOv3 e o TVIP (CM, $r=0,356; p < 0,001$; SL, $r = 0,490; p < 0,001$; OLV, $r=0,487; p < 0,001$; IH, $r=0,331; < 0,001$; CF, $r=0,411;$

$p < 0,001$; CH, $r = 0,300$; $p < 0,001$; Total, $r = 0,563$; $p < 0,001$). Constatou-se o aumento do desempenho em compreensão da linguagem oral função da progressão da idade ($F_{\text{Total}} [7, 317] = 14,059$; $p < 0,001$) e do ano escolar ($F_{\text{Total}} [4, 320] = 24,554$; $p < 0,001$). Foram encontradas diferenças estatisticamente significativas no desempenho da prova CM ($t[313] = 2,094$; $p = 0,037$) e no total da BILOv3 em função do gênero ($t[318] = 1,970$; $p = 0,050$), indicando que as meninas obtiveram médias mais altas ($M = 105,27$; $DP = 13,05$) do que os meninos ($M = 102,20$; $DP = 15,07$). A análise do tempo da BILOv3 indicou o decréscimo de tempo à medida da progressão da idade ($F_{\text{Total}} [7, 317] = 2,558$; $p = 0,014$) e do ano escolar ($F_{\text{Total}} [4, 320] = 5,158$; $p < 0,001$). Por meio do Alfa de Cronbach foram evidenciados valores de consistência interna para as provas (CM = 0,71; SL = 0,85; OLV = 0,86; IH = 0,76; CF = 0,55 e CH = 0,54), sendo o efeito somado de todas as provas igual a 0,91. Tendo em vista os resultados encontrados, a pesquisadora, constatou adequadas características psicométricas para a BILOv3.

A partir dos estudos descritos observou-se que a BILOv3 já conta com evidências de validade, contudo considera-se que novos estudos devam ser feitos, ampliando a amostra, principalmente com pré-escolares, e assegurando que a BILO corresponda às qualidades psicométricas necessárias para que possa ser utilizada como um teste psicológico.

1.5 Estudos com construtos correlatos

No presente estudo, destaque é dado para a participação do vocabulário receptivo e da inteligência como requisitos para a efetiva compreensão da linguagem oral. É possível observar, a partir do Modelo Neuropsicolinguístico, a presença de tais requisitos no que se refere ao léxico e às operações mentais – percepção, atenção, memória e função executiva -

que são necessárias para relacionar os processos que levam à competência linguística e comunicativa.

Strauss, Sherman e Spreen (2006) descrevendo o *Peabody Picture Vocabulary Test Third Edition* (PPVT III), reforçam a forte correlação entre o escore de vocabulário os quocientes de inteligência (QIs). De acordo com Eysenck (1994), o vocabulário receptivo correlaciona-se fortemente com quociente de inteligência, pois a capacidade de usar informações contextuais e frasais para inferir o significado das palavras desconhecidas é um dos principais meios de adquirir vocabulário. Deste modo, considera-se que a aquisição do vocabulário receptivo auditivo, assim como a capacidade de compreensão da linguagem oral, estão diretamente relacionadas ao desenvolvimento da habilidade inferencial.

A relação entre compreensão da linguagem oral e vocabulário e com inteligência já foi explorada em estudos com a BILO (Joly & Almeida, 2008; Joly, Martins, Lopes & Lemos, 2009; Istome, 2010; e Cruz, 2012), conforme anteriormente exposto. Enfatiza-se que tais estudos demonstraram que o vocabulário receptivo, avaliado pelo Teste de Vocabulário por Imagens Peabody – TVIP, correlaciona-se positiva e significativamente com os escores por provas e total da BILO, bem como a análise de regressão múltipla realizada no estudo de Joly e Piovezan (2012) apontou o raciocínio geral, avaliado pela escala de Maturidade Mental Colúmbia (Alves & Duarte, 1994), como representante de 29% das habilidades necessárias para um bom desempenho na BILOv1. A partir destes estudos, para fins de evidenciar validade para a terceira versão da BILO, o presente estudo buscou correlações positivas entre linguagem oral, vocabulário e inteligência. Foram medidas de critério correlacionadas com os escores da BILO os escores do Testes de Vocabulário por Imagens Peabody, computadorizado (Macedo, Capovilla, Duduchi, D'Antino & Firmo, 2006) e das Matrizes Progressivas Coloridas de Raven (Angelini, Alves, Custódio, Duarte & Duarte, 1992).

O TVIP tem como objetivo avaliar o desenvolvimento lexical no domínio receptivo auditivo, isto é, a compreensão do significado das palavras ouvidas no que se refere às pessoas, ações, qualidades, partes do corpo, tempo, natureza, lugares, objetos, animais, termos matemáticos, ferramentas e instrumentos (Capovilla & Capovilla, 1997). Considera-se que o léxico é um requisito importante para a competência linguística, descrita por Kaufman (1996), no que se refere às regras semânticas, conforme já mencionado.

Conforme exposto por Capovilla, Nunes, Araújo, Nogueira e Bernat (1997), o TVIP é um instrumento concebido originalmente por Dunn em 1959, revisado por Dunn e Dunn em 1981 e adaptado para o espanhol por Dunn, Padilla, Lugo e Dunn em 1986. A versão brasileira foi traduzida e adaptada da versão hispano-americana por Capovilla e Capovilla (1997), e possui evidências de validade encontradas por estes mesmos pesquisadores para a população de 2 anos e 6 meses a 18 anos. A versão computadorizada do TVIP, que será utilizada no presente estudo, foi desenvolvida por Macedo, Capovilla; Duduchi, D'Antino e Firmo (2006), e tem como vantagens demandar menos tempo de aplicação, facilidade na tabulação de dados e aferição de resultados.

O TVIP computadorizado foi utilizado por Macedo, Capovilla, Duduchi, D'Antino e Firmo (2006) para avaliar a linguagem receptiva em crianças pré-escolares, comparando as pontuações dos pré-escolares nas versões tradicional e computadorizada do teste. Foram avaliadas 90 crianças de uma escola particular do município de São Paulo, sendo 30 crianças do infantil I (15 meninos e 15 meninas) com idade média de 4 anos; 30 do infantil II (17 meninos e 13 meninas) com idade média de 5 anos e 30 do infantil III (18 meninos e 12 meninas), todas com autorização formal dos pais. Foram aplicadas coletivamente a versão tradicional e a computadorizada do TVIP em duas sessões separadas por uma semana.

Na primeira sessão, metade das crianças foi submetida à versão tradicional, e a outra metade, à versão computadorizada. Na segunda avaliação, uma semana depois, a primeira metade foi submetida à versão computadorizada, e a segunda metade, à tradicional. A análise dos resultados revelou que, em média, as crianças acertaram 13 itens a mais na versão tradicional do que na computadorizada ($t = 11,331$; $p = 0,000$) e que levaram em média 6 minutos a menos na versão computadorizada ($t = 4,926$; $p = 0,000$). Embora tenha sido observada discrepância na pontuação entre as duas versões, a análise de regressão do escore na versão tradicional em relação à versão computadorizada revelou correlação positiva ($r = 0,646$; $r^2 = 0,418$) significativa entre os desempenhos nas duas versões ($F[1,89] = 63,187$; $p = 0,000$).

A análise de variância (ANOVA) do efeito da idade sobre o escore no teste TVIP indicou efeito significativo para a versão tradicional ($F[2,89] = 24,878$; $p = 0,000$) e computadorizada ($F[2,89] = 11,890$; $p = 0,000$), identificando-se que as crianças de 6 anos acertaram mais do que as demais nas duas versões, e as de 5 anos acertaram mais do que as de 4 anos na versão tradicional. A Análise de variância (ANOVA) do efeito da idade sobre o tempo para realização do teste TVIP revelou efeito significativo para a versão tradicional ($F[2,89] = 4,868$; $p = 0,010$), mas não a computadorizada, denotando que as crianças com 6 anos foram mais rápidas do que as com 4 anos, e não houve diferença entre as crianças com 5 anos e as demais. O Teste t para medidas independentes não revelou efeito de sexo para tempo e pontuação dos testes TVIP nas versões tradicional e computadorizada. Não foi observado efeito da ordem de aplicação dos testes sobre o tempo de execução ou dos escores obtidos nas duas versões. Análise de frequência de acertos na versão computadorizada indica alta similaridade com as da versão tradicional. Os itens mais difíceis foram: ler, canguru, elipse, fragmento, ascender, jubilosa e perpendicular.

Em relação ao uso das Matrizes Progressivas Coloridas de Raven (Angelini, Alves,

Custódio, Duarte e Duarte, 1992) como medida de inteligência a ser utilizada na correlação com a BILO e com o TVIP, esta se justifica por se propor a medir os processos intelectuais de crianças na faixa de 5 a 11 anos. Mais precisamente, este instrumento tem como foco avaliativo a capacidade *edutiva* para relações e correlatos, processo este componente do fator *g* de inteligência descrito por Spearman na década de vinte (Angelini e cols., 1992). Tendo em vista que este instrumento é baseado na teoria de Spearman (1927), cabe, para fins de compreensão da concepção de inteligência a ser utilizada neste estudo, especificar resumidamente alguns pontos principais de sua Teoria Eclética dos dois fatores, ou Bifatorial.

A partir de experiências fatoriais, Spearman (1927) demonstrou que a inteligência era composta por um fator geral (*g*) – “fator quantitativo único da inteligência comum e fundamental para todas as funções cognitivas de um mesmo indivíduo” (p. 14); um fator específico (*e*) “próprio de cada habilidade particular que não depende nem se correlaciona com *g* nem com outro *e*” (p. 14) – e um fator de grupo “comum a muitas habilidades de um conjunto afim; relaciona unitariamente grande parte de um conjunto dado de habilidades” (p. 15). Em termos de funcionamento cognitivo, Spearman definiu o fator *g* composto por três leis fundamentais, quais sejam, a Apreensão da própria experiência – que expressa a capacidade introspectiva de observação de suas próprias sensações e percepções –; a Educação de relações – referente à capacidade de estabelecer relações entre duas ou mais idéias percebidas ou pensadas, categorizando-as em classes – e a Educação de correlatos, que alude à capacidade de incorporar ao pensamento novas idéias a partir da correlação com idéias e relações que já possuía.

Levando em consideração que o fator *g* se refere a um componente comum subjacente a diversas dimensões cognitivas avaliadas pelos testes, é de se esperar que sejam observadas correlações entre o escore de capacidade edutiva com os escores das provas da

BILO, bem como com o do vocabulário. Além disso, diante da capacidade de eduzir relações, admite-se que a aquisição da compreensão da linguagem envolve, entre outros, o uso de relações contextuais para entender o conteúdo transmitido, tanto na competência linguística como na competência comunicativa (Eysenck, 1994; Kaufman, 1996,).

Trevisan, Montiel, Dias, e Capovilla (2008), encontraram a relação entre o vocabulário e a inteligência não verbal, quando avaliaram a habilidade de vocabulário receptivo em crianças de 1^a à 4^a série do ensino fundamental. Foram avaliadas 363 crianças de ambos os sexos com idades entre 6 anos e 6 meses e 11 anos e 9 meses de uma escola municipal do interior de São Paulo. Foram utilizados como instrumentos as Matrizes Progressivas de Raven e o Teste de vocabulário por Imagens Peabody (TVIP) versão computadorizada. As avaliações ocorreram em duas sessões. Foi realizada uma Análise de Covariância (ANCOVA) tendo o percentil do Raven como covariante e a idade e série como fatores, de modo a verificar o efeito da idade e da série sobre os escores do TVIP com o controle sobre efeito da inteligência não-verbal. A ANCOVA revelou efeito significativo da idade ($F[3, 251] = 9,05; p \leq 0,000$) e do percentil no Raven ($F[1, 249] = 21,38; p \leq 0,000$). Também revelou o efeito significativo da série ($F[3, 251] = 20,42; p \leq 0,000$) e do percentil no Raven ($F[1, 251] = 17,92; p \leq 0,000$). Observou-se, pois, que os escores aumentaram com a progressão da idade e da série e a habilidade de vocabulário receptivo esteve relacionada com a inteligência não verbal, com a idade e com a série escolar. Além disso, foi encontrado um bom índice de precisão do TVIP, verificado pelo alpha de Crombach igual a 0,87.

Com base no exposto a seguir são especificados de modo sistemático o objetivo geral e os objetivos específicos do presente estudo.

2. OBJETIVO

Tendo em vista a incidência de transtornos da linguagem na infância (entre 1 e 12%), sendo que nas crianças pré-escolares a média é de 5%, e a ligação destes transtornos com algum atraso no processamento cognitivo ou distúrbio de aprendizagem por volta dos nove anos de idade (Pedroso & Rotta, 2006), faz-se indispensável caracterizar nas fases iniciais de escolarização o desempenho dos alunos em linguagem oral. Para tanto, considera-se de suma importância o desenvolvimento de instrumentos que auxiliem os profissionais a caracterizarem o desempenho e identificar os aspectos deficitários do avaliando, para assim direcionar condutas terapêuticas adequadas (Chevrie-Muller & Narbona, 2005). Tais instrumentos ainda são escassos no contexto brasileiro (Capovilla e cols., 2006, Gurgel, Plentz, Joly & Reppold, 2010).

Ao lado disso, tem-se que, para o desenvolvimento e uso de instrumentos de testagem psicológica, sejam tomados cuidados metodológicos que assegurem que o teste possua qualidades psicométricas que lhe confirmem evidências de validade, de precisão e normatização (Anastasi & Urbina, 2000). Isto posto, o presente estudo buscou evidenciar as qualidades psicométricas de um instrumento informatizado destinado a avaliar a compreensão da linguagem oral, tendo como base a relação deste com outras variáveis, considerando, ainda, o histórico de desenvolvimento do instrumento. Os objetivos específicos estão descritos a seguir.

- a) Buscar evidências de validade com base na relação com o desempenho em vocabulário e com a inteligência;
- b) Verificar diferenças relativas à série e à idade que lhe forneçam evidências de validade com base na relação no critério desenvolvimental e de aprendizagem.

3. MÉTODO

3.1 Participantes

Participaram do presente estudo 262 alunos, com idades que variavam de cinco a oito anos ($M = 6,37$ e $DP = 0,89$), regularmente matriculados no último ano da Educação Infantil ($N = 85$) e primeiro ($N = 96$) e segundo ($N = 81$) anos do Ensino Fundamental da rede de ensino público de uma cidade do interior do estado de São Paulo com 13.105 habitantes.

A proporção de alunos que participaram da pesquisa em relação ao número de alunos matriculados no mesmo ano escolar no município todo, corresponde a 52,15% dos alunos do último ano do Ensino Infantil, 57,83% dos alunos do 1º ano e 58,27% dos alunos do 2º ano do Ensino Fundamental. As escolas que participaram da pesquisa localizam-se na zona urbana, mas recebem alunos da zona rural. O último Índice de Desenvolvimento da Educação Básica – IDEB (2011) do município foi de 4,8, apresentando uma queda de 4% em relação à última avaliação (2009) e estando 16% abaixo da meta projetada que é 5,7.

Na escola de Ensino Infantil em que foi desenvolvida a pesquisa não havia sala de informática e a grande maioria dos alunos nunca havia tido contato com o computador. Já a escola de Ensino Fundamental, contava com sala de informática composta por 14 computadores, conectados à internet, onde os alunos faziam aulas semanais de informática.

Dos 262 alunos que participaram, 114 eram meninas (43,51%) e 148 meninos (56,49%). A Figura 1 apresenta a distribuição dos alunos em função do sexo por ano escolar frequentado.

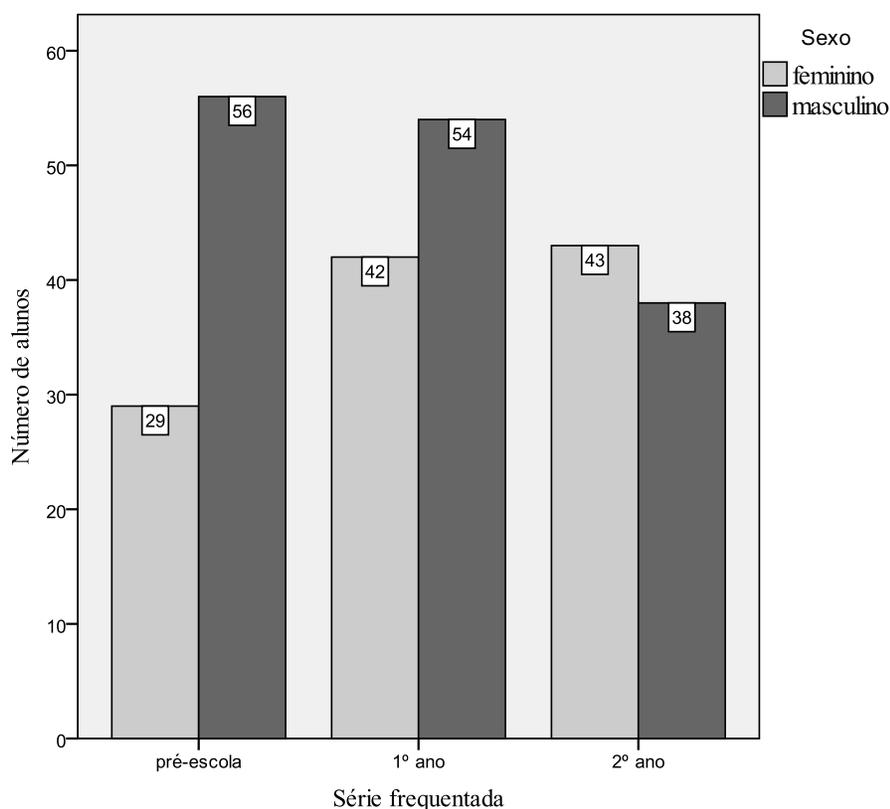


Figura 2. Distribuição do número de alunos por sexo em cada ano escolar frequentado.

Quando observada a distribuição de meninos e meninas por ano escolar frequentado (Figura 2) nota-se que há maior diferença no número de alunos por sexo no Ensino Infantil, em que de 85 alunos, 65,88% são do sexo masculino. Somente no 2º ano há prevalência do sexo feminino sobre o masculino, contudo esta é a série em que o sexo está mais equitativamente distribuído.

Com relação à distribuição das idades dos participantes segundo o ano escolar frequentado (Figura 3), observa-se que prevalece a idade de 5 anos para o último ano da Educação Infantil ($M = 5,45$; $DP = 0,50$), de 6 anos para 1º ano ($M = 6,40$; $DP = 0,49$) e de 7 anos para o 2º ano ($M = 7,31$; $DP = 0,49$) do Ensino Fundamental indicando adequação à série frequentada.

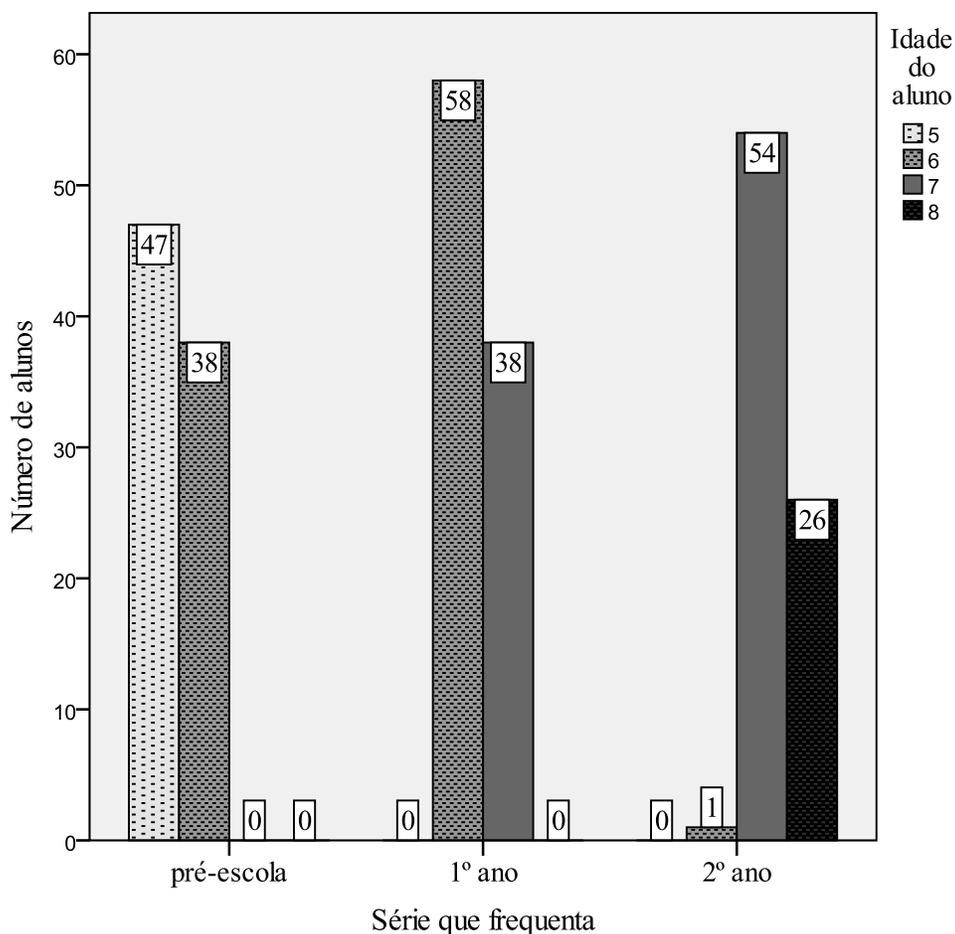


Figura 3. Distribuição da idade dos alunos por nível escolar frequentado.

3.2. Instrumentos e Materiais

3.2.1 Bateria Informatizada de Linguagem Oral-BILOv3 (Joly, 2009)

A BILO foi construída com base na revisão da literatura (American Speech-Language-Hearing Association, 2007; Ellis, 1995; Klein, 2004; Paris & Sthal, 2005; Tellez, 2004, conforme especificado por Joly e cols, 2008) e na análise de testes equivalentes publicados no exterior (DRP, International Dyslexia Test, conforme especificado por Joly e cols, 2008). Tem por objetivo avaliar a compreensão da linguagem oral em alunos da

educação infantil ao quinto ano do ensino fundamental. Sendo um instrumento informatizado, foi desenvolvido por meio de software que tem recursos de multimídia e interface com banco de dados para armazenamento das respostas dadas. A instalação do programa é feita automaticamente ao se inserir o CD no computador. A BILO é instalada em C:\BILO (disco local) e um ícone na área de trabalho é gerado. Conforme os respondentes realizam as provas, os dados gerados pelas respostas são gravados automaticamente em C:\BILO\DADOS, em arquivos denominados “banco”, “cm”, “sl”, “olv”, “cf”, “ch” e “ih” com extensão “.txt”, que podem ser abertos com editor de textos, e com extensão “.xls”, que podem ser abertos em uma planilha de dados.

Em se tratando de uma avaliação voltada para a linguagem oral com respondentes que não adquiriram o código formal de leitura, optou-se por respostas apresentadas como desenhos feitos apenas com linhas, em preto e branco. Estes foram selecionados considerando-se símbolos, significados e contextos familiares à criança. As instruções gerais e específicas de cada prova e item são apresentadas oralmente pelo programa por fones de ouvido e aparecem também impressas na tela usando letra bastão. Cada tela tem apenas um item da prova; após a escolha da opção, o respondente é solicitado a confirmar sua resposta para continuar a prova. Em um banco de dados são registrados a opção de resposta e o tempo total por item.

Há um tutorial interativo indicando quais os recursos de interface do computador são utilizados para responder à BILO. Nesse tutorial, é solicitado que o respondente complete alguns itens visando exercitar tais habilidades necessárias, como “clique”, “arrastar” a figura e selecionar o item como resposta.

A BILO é constituída por seis provas, com 10 itens cada, que avaliam construtos específicos relacionados à linguagem oral, conforme segue:

CM – Compreensão Morfossintática: tem por objetivo avaliar a relação entre um vocábulo e a sua representação gráfica, revelando a compreensão do significado. Cada item da prova formado por duas orações com um termo integrante que pode ser uma logomarca, um substantivo, adjetivo, verbo ou advérbio, a ser compreendido pelo respondente. A tarefa do respondente é escolher, dentre as três opções, a que revela a relação correta entre desenho e nome.

CM EX2. SOU MUITO CARINHOSO E FOFINHO. SOU O GATO JUJU.

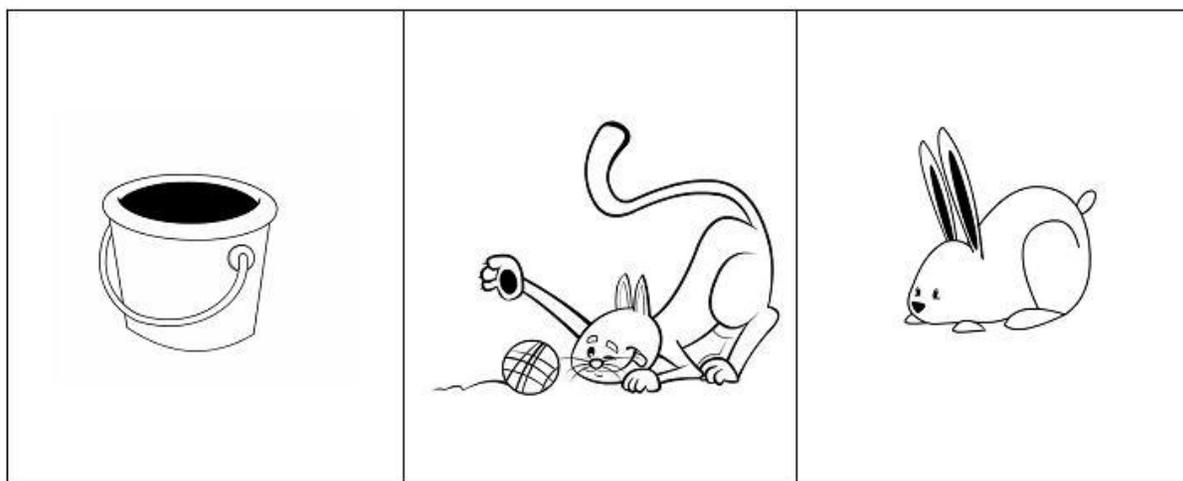


Figura 4. Exemplo 2 prova CM.

SL – Sequencia Lógica: Visa avaliar a organização lógica e temporal do respondente para estímulos visuais organizados em cenas cujo conjunto seqüenciado resulta em uma história. Cada item é formado por uma série de cenas que, organizadas em uma sequencia com começo, meio e fim, adquirem significado de uma história. A tarefa do respondente é organizar as cenas de cada série de modo que haja relação de causa e efeito e contigüidade entre elas. A criança orienta-se exclusivamente pelos estímulos visuais. Trata-se da única prova da BILO que não agrega estímulos orais às imagens para sua execução.

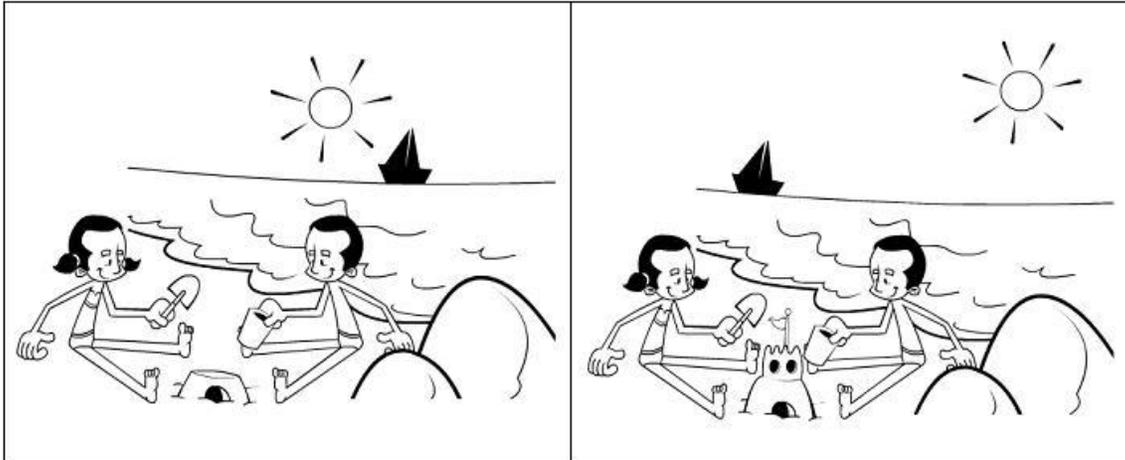
SL EX1. Praia

Figura 5. Exemplo 1 prova SL.

OLV – Organização Lógico verbal: Avalia a compreensão receptiva do conteúdo de uma história apresentada por completo em vídeo e oralmente, por meio de uma organização lógica de cenas que representam esse conteúdo. Cada item é composto por uma história e por uma série de cenas. Nessa tarefa, o respondente ouve a história e deve organizar as cenas de cada série, de acordo com a sequencia lógica do conteúdo visual e verbal.

OLV EX1. O MENINO GANHOU UMA PIPA. ELE EMPINOU NO PARQUE.

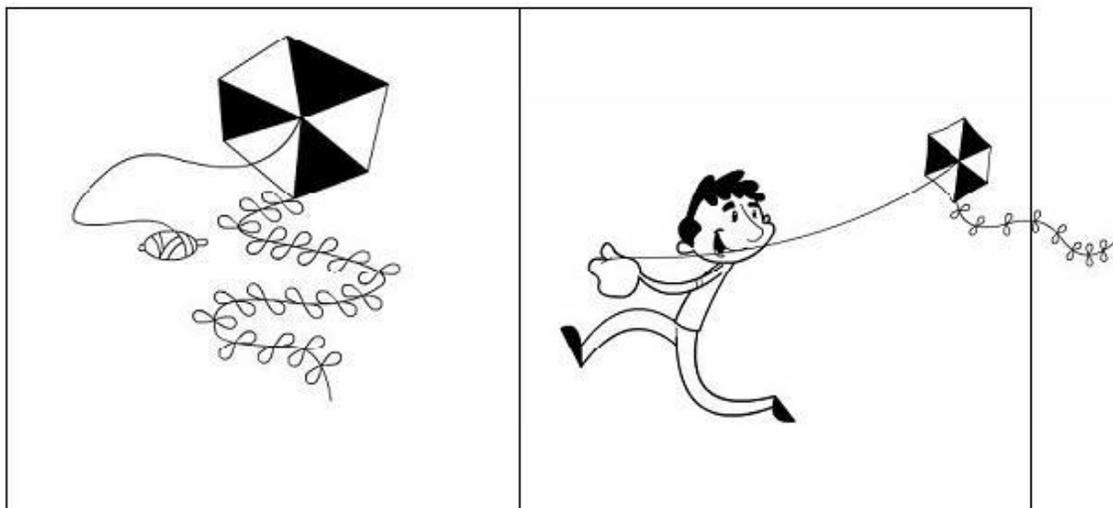


Figura 6. Exemplo 1 prova OLV.

IH – Interpretação de Histórias: Avalia a compreensão receptiva da história por meio de questões de múltipla escolha. Todos os itens referem-se à compreensão literal da história. A tarefa do respondente é assistir ao áudio-vídeo (que só pode visto duas vezes) e responder às posteriores questões, escolhendo uma opção (dentre três) de resposta por item.

O GATO DE MARIANA

EU TENHO UMA GATA. SEU NOME É MIMI. GANHEI MIMI DO MEU AVÔ. MIMI É MUITO BRINCALHONA. ELA É A MINHA MELHOR AMIGA. ADORO MIMI

IH1. QUAL É O BICHO DE MARIANA?

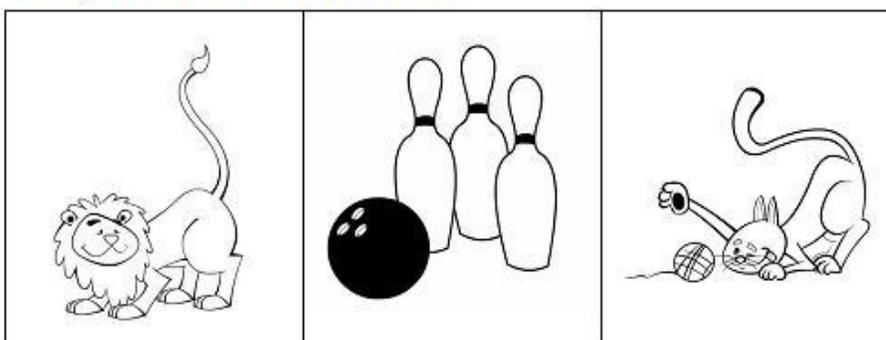


Figura 7. Item 1 prova IH.

CF – Completar Frases: Avalia a compreensão de palavras organizadas em frases. Deve-se destacar que um dos substantivos é omitido de acordo com a organização textual baseada no Sistema Orientado de Cloze (Joly, 2007). Esse sistema possibilita a organização de um texto para avaliar compreensão em leitura, com base em critérios específicos relativos a número de palavras, omissões de vocábulos, tamanho de lacunas e opções de resposta, visando determinar níveis de dificuldade de texto em função da habilidade de compreensão dos respondentes. As opções de resposta por item contam com uma alternativa com a mesma classificação gramatical da resposta correta e a outra com uma palavra qualquer, desde que seja classificada como diferente das outras duas opções.

CF 2 O _____ COME CENOURA.

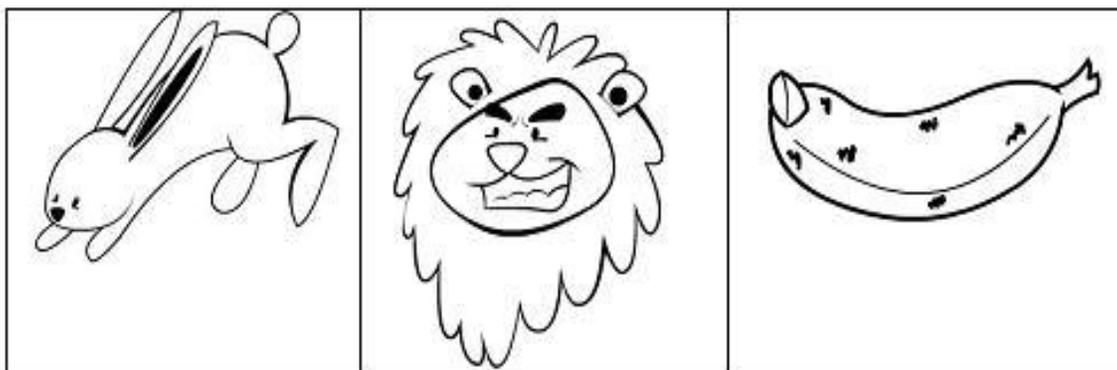
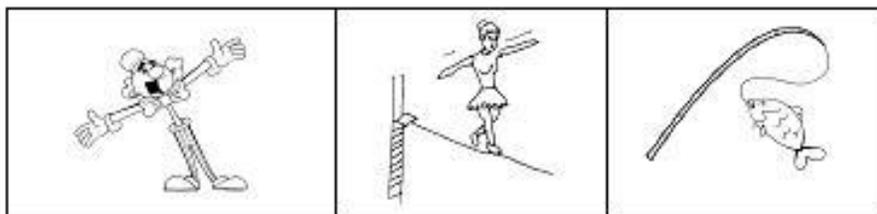


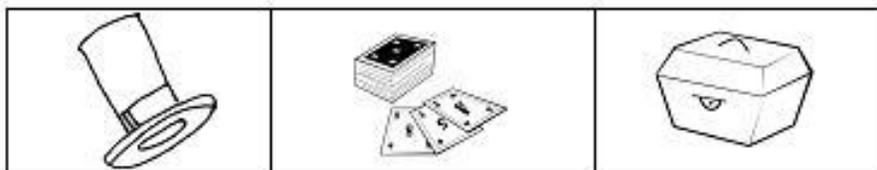
Figura 8. Item 2 prova CF.

CH – Completar Histórias: Avalia a compreensão oral de histórias. Cada história completa é, primeiro, apresentada em forma de vídeo (imagens com áudio) para, em seguida, ser visualizada na tela na forma impressa, acompanhada do respectivo áudio com as opções de resposta (figuras) por item. A tarefa do respondente é assistir ao vídeo e ouvir a história que é contada duas vezes e escolher uma opção por item que melhor dê sentido a palavra omitida. Também foi baseada no Sistema Orientado de Cloze (Joly, 2007).

O CIRCO
 CH1. O CIRCO É MUITO DIVERTIDO. TEM A S BRINCADEIRA S E CAMBALHOTA S DO _____.



CH2. TEM O MÁ GICO QUE FAZ TRUQUES COM A S CARTA S DE _____.



CH3. OS BICHOS SÃO AINDA MAIS ENGRAÇADOS. O _____ ENCHE A TROMBA DE ÁGUA.



COM UM JATO MOLHA A JUBA DO FERROZ LEÃO. ELE FICA MUITO BRAVO.

Figura 9. Itens 1, 2 e 3 prova CH.

3.2.2. *Teste de Vocabulário por Imagens Peabody – TVIP Versão computadorizada*
 (Macedo; Capovilla; Duduchi; D'Antino & Firmo, 2006)

Avalia o desenvolvimento lexical no domínio receptivo em pessoas com idades entre 2 anos e 6 meses até 18 anos. Permite uma avaliação objetiva, rápida e precisa do vocabulário receptivo auditivo em ampla variedade de áreas, incluindo pessoas, ações, qualidades, partes do corpo, tempo, natureza, lugares, objetos, animais, termos matemáticos, ferramentas e instrumentos. É utilizado para indicar o nível de desenvolvimento da linguagem receptiva em pré-escolares, bem como em crianças e

adultos incapazes de ler, escrever ou mesmo de falar, já que, a figura corresponde à palavra falada pelo examinador. A criança deve apontar a figura que corresponde à palavra que ouve e marcar um x.

No presente estudo será usada a versão computadorizada do TVIP (Macedo; Capovilla; Duduchi; D'Antino & Firmo, 2006) que foi produzida a partir da adaptação hispano-americana de 125 itens, cada um com quatro desenhos, adaptada e validada e normatizada no Brasil para a faixa etária de 2 a 14 anos (Dunn, Dunn, Capovilla & Capovilla, no prelo). As mensagens solicitando que a criança escolha uma das figuras são emitidas a partir de arquivos de sons digitalizados e salvos no formato WAV. Uma vantagem da versão computadorizada do teste em relação à versão tradicional é a de registrar todas as respostas do avaliando, bem como o tempo consumido em responder.

Macedo, Capovilla, Duduchi, D'Antino e Firmo (2006) comparando as pontuações de 90 pré-escolares - com idades entre quatro e seis anos - nas versões tradicional e computadorizada do teste, verificaram que, embora tenha sido observada discrepância na pontuação entre as duas versões (mais acertos na versão lápis e papel), a análise de regressão do escore na versão tradicional em relação à versão computadorizada revelou correlação positiva ($r = 0,646$; $r^2 = 0,418$) significativa entre os desempenhos nas duas versões ($F[1, 89] = 63,187$; $p \leq 0,001$), indicando evidência de validade para a versão computadorizada. Também foi constatado efeito da idade sobre o escore no teste TVIP computadorizado ($F[2, 89] = 11,890$; $p \leq 0,001$), identificando-se que as crianças de 6 anos acertaram mais do que as demais nas duas versões, atribuindo assim, validade teste-critério. O estudo de Tevisan, Montiel e Dias (2008) revelou que o TVIP computadorizado possui Alpha de Cronbach igual a 0,87.

Exemplo:

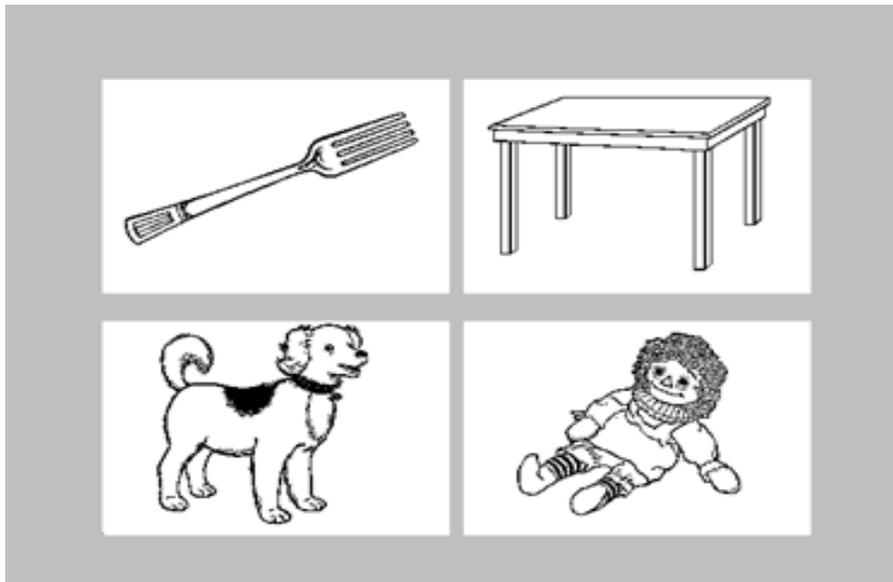


Figura 10. Exemplo de item do TVIP.

Instrução audio-verbal: *Clique na figura que o computador disser o nome: boneca.*

3.2.3. *Matrizes Progressivas Coloridas de Raven – Escala Especial (Angelini, Alves, Custódio, Duarte e Duarte, 1999)*

É um teste que avalia a capacidade de estabelecer relações, ou seja, de estabelecer relações entre variáveis, o que denota na *capacidade de desenvolver construtos de nível superior, que tornam mais fácil pensar sobre situações e eventos complexos* (Angelini, Alves, Custódio, Duarte e Duarte, 1999, p.6). É destinado a avaliar tal capacidade em crianças entre 4 anos e nove meses e 11 anos e 8 meses, pessoas com deficiência e idosos. Este teste é derivado do teste Matrizes Progressivas Standart desenvolvido por J. Raven para avaliar sujeitos de 12 a 65 anos. Raven ao conceber o instrumento seguiu os princípios de três vertentes teóricas, quais sejam, a teoria dos dois fatores de Spearman (procedimentos estatísticos que demonstravam que as relações entre uma série de variáveis

podiam ser explicadas por um único grande fator, o fator geral ou g.); a teoria da Gestalt (uso itens de caráter gráfico para evitar influências de elementos culturais e de treinamento) e a teoria do desenvolvimento cognitivo (a habilidade se desenvolve linearmente por etapas com a maturação do organismo).

O teste Matrizes Progressivas Coloridas de Raven, foi publicado, inicialmente, na Inglaterra em 1947, sendo que houve uma revisão dos itens em 1956. A normatização e padronização brasileiras, por Angelini, Alves, Custódio, Duarte e Duarte (1999) derivam da revisão de 1956.

O teste apresenta três séries (A, Ab e B) com 12 problemas cada, que são destinadas a avaliar com maior precisão os processos intelectuais de crianças. O sucesso na série A depende da capacidade do indivíduo para completar padrões, na série Ab depende da capacidade para perceber diferenças discretas em figuras e na série B depende da capacidade do indivíduo em fazer analogias. Estes itens representam figuras nas quais falta uma parte que deve ser completada, utilizando um dos seis encaixes apresentados como alternativas de solução. Para completar corretamente a figura, a criança tem que solucionar o problema que a figura implica.

O teste foi padronizado para população brasileira por Angelini, Alves, Custódio, Duarte e Duarte (1999). Compuseram a amostra de padronização brasileira 1547 crianças da pré-escola aos seis anos iniciais do Ensino Fundamental de escolas públicas e particulares da cidade de São Paulo. Os pesquisadores constataram evidências de validade utilizando a idade como critério e a consistência interna do instrumento, além disso, a precisão foi verificada por meio da técnica das duas metades. Bandeira, Alves, Giacomel e Lorenzatto (2004), estabeleceram normas em percentis por faixa etária para as crianças de Porto Alegre – RS, avaliando 779 crianças matriculadas em escolas estaduais, com idades entre 4 e 11 anos.

Exemplo:

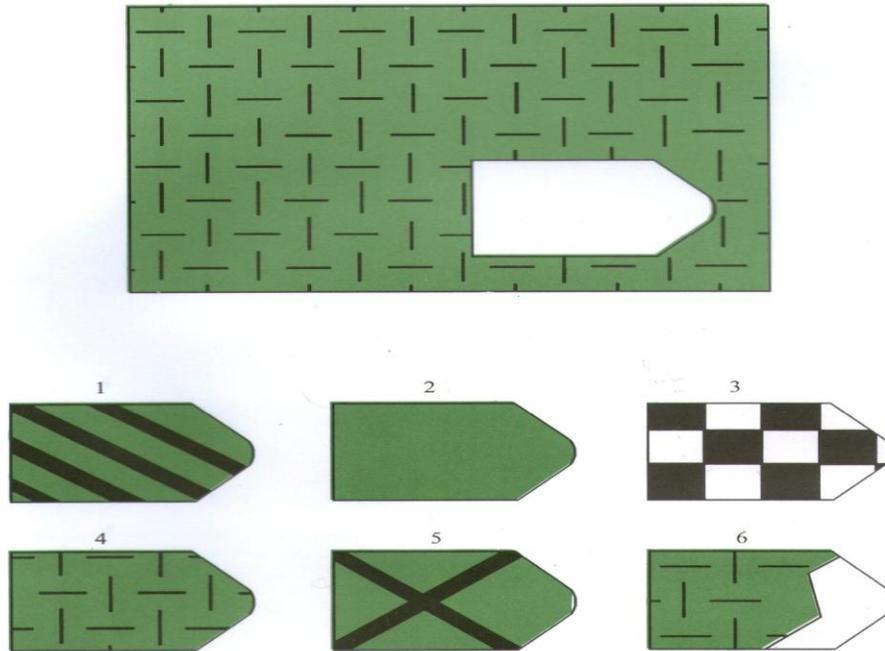


Figura 11. Exemplo de item Matrizes Progressivas Coloridas de Raven – Escala Especial.

3.3. Equipamentos

Microcomputadores que possuam como requisitos mínimos do sistema processador Pentium 4; memória mínima de 256 MB; espaço em disco de 180 MB; plataforma Windows XP, Vista ou 7; com resolução de tela 1024 por 768 pixels; e fones de ouvido.

3.4. Procedimentos

- Procedimento de Aplicação

Inicialmente foi solicitada a autorização formal da Secretaria Municipal de Educação. Diante do consentimento da autoridade responsável, o projeto de pesquisa foi submetido à avaliação do Comitê de Ética da Universidade São Francisco, sendo aprovado sob Protocolo CAAE 0399.0142.000-11. Desta feita, foram contatados os diretores das

escolas participantes para apresentação do projeto e confecção do cronograma de aplicações. Foi realizada uma reunião com os professores para a apresentação do projeto e apresentação do cronograma. Foi enviado o Termo de Consentimento Livre e Esclarecido para os responsáveis, explicando a pesquisa e solicitando autorização para a participação do filho. Somente três alunos não participaram por não apresentarem o consentimento dos pais, indicando que houve interesse e colaboração da população para a realização da pesquisa.

A aplicação ocorreu em três sessões, uma para cada instrumento, em dias diferentes na mesma semana ou na semana seguinte. Foi aplicada a BILO na primeira sessão, as Matrizes Progressivas Coloridas de Raven na segunda sessão e o TVIP na terceira sessão. Houve o cuidado de aplicar a BILO_{v3} na primeira sessão, antes do TVIP, porque a bateria possui um tutorial de treinamento das habilidades de manuseio do mouse, garantindo que os alunos que não possuíam contato com a informática pudessem aprender a utilizar os recursos antes de iniciarem as provas da BILO_{v3} e, posteriormente, realizarem o TVIP. Exceto pelo zelo em aplicar a BILO_{v3} antes do TVIP, a sequência estabelecida para a aplicação foi decorrente da adequação da pesquisa ao cronograma de atividades das escolas, levando-se em conta a disponibilidade de liberação dos alunos e dos locais de aplicação dos testes.

Nos anos do Ensino Fundamental as aplicações dos instrumentos informatizados foram realizadas coletivamente no laboratório de informática da escola, nos 14 computadores da escola e nos 6 computadores portáteis levados pelos aplicadores. Já na escola do Ensino Infantil, onde não há laboratório de informática, as aplicações ocorreram na biblioteca, em 5 computadores portáteis levados pelos avaliadores. As aplicações das Matrizes Progressivas Coloridas de Raven foram realizadas individualmente, em sala de aula disponibilizada pela escola. As aplicações foram realizadas de acordo com as

instruções específicas de cada instrumento. O tempo de aplicação da BILOv3 foi de aproximadamente 40 minutos, do TVIP e do Raven em 10 minutos cada.

As aplicações da BILOv3 e do TVIP foram realizadas pela pesquisadora responsável, por um aluno de mestrado e um aluno de doutorado em avaliação psicológica, ambos psicólogos. As aplicações das Matrizes Progressivas Coloridas de Raven, foram conduzidas pela pesquisadora responsável, por um aluno de mestrado, dois alunos de doutorado em Avaliação Psicológica e dois bolsistas de Iniciação Científica, todos com treinamento prévio.

- Procedimento de correção

Os critérios de correção da BILO são por item, por prova e geral. Cada prova contém 10 itens de múltipla escolha com três opções de resposta. Duas opções referem-se a conteúdos pertencentes à mesma classe gramatical ou categoria, sendo um deles a resposta correta (que vale 2 pontos) e a outra opção uma resposta semelhante (que vale 1 ponto). A terceira opção é uma resposta de outra classe gramatical ou categoria diferente da opção correta é pontuada com zero. A somatória dos pontos de todos os itens de cada prova determina o escore por prova, sendo que o escore máximo obtido é de 20 pontos. A somatória dos escores das seis provas resulta no escore total da BILO que pode chegar até 120 pontos.

O TVIP computadorizado gera um arquivo para cada protocolo, sendo que, os itens respondidos corretamente recebem “C” e os errados “E”, estes posteriormente são convertidos em 1 e 0, respectivamente. A soma de todos os itens resulta no escore total do TVIP que pode alcançar 125 pontos. Já para a correção dos protocolos das Matrizes Progressivas Coloridas de Raven, atribui-se um ponto para cada acerto, obtendo-se um escore bruto do total de acertos (pontuação máxima igual a 36) que foi transformado em

percentil de acordo com o grupo de referência que se está estudando, neste caso, alunos de escola pública. Os dados do percentil são analisados seguindo os escores de padronização. Deste modo, é possível classificar o indivíduo em relação à população normal, em “intelectualmente deficiente” (resultado igual ou inferior a 5), “definidamente abaixo da média na capacidade mental” (resultado de 6 a 25), “intelectualmente médio” (de 26 a 74), “definidamente acima da média” (de 75 a 94) e “intelectualmente superior” (igual ou superior a 95).

4. RESULTADOS E DISCUSSÃO

Tendo o presente estudo o objetivo de buscar evidências de validade para a Bateria Informatizada de Linguagem Oral versão 3 (BILOv3), por meio da correlação com o desempenho em vocabulário e com a inteligência, bem como verificar se há diferenças no desempenho em função do ano escolar, idade e gênero, inicialmente serão apresentados os dados relativos às análises estatísticas descritivas e inferenciais. Estão, portanto, descritas as análises do desempenho no Teste de Vocabulário por Imagens *Peabody* – (TVIP), nas Matrizes Progressivas Coloridas de Raven (Raven) e na BILOv3, respectivamente. Tendo, a cada início de descritiva, as análises exploratórias do teste em questão.

Após a descrição dos desempenhos, serão expostas as análises do tempo utilizado para a execução da BILOv3 e as possíveis diferenças relativas às características dos alunos. Por fim, será apresentada a análise de correlação entre os escores dos três instrumentos e as diferenças relativas aos desempenhos por grupos extremos.

4.1 *Teste de Vocabulário por Imagens Peabody – (TVIP)*

São apresentadas, na sequência, a análise exploratória, as análises descritivas e as inferenciais do escore e do tempo de execução no TVIP. Os dados referentes à análise exploratória do desempenho em vocabulário receptivo são apresentados na Tabela 1, que indica que os participantes obtiveram notas que variaram de 23 a 87, sendo que a média foi de 61,95 ($DP = 11,14$). É possível afirmar, com 95% de confiança, que o intervalo da média está entre 60,60 e 63,31. A mediana indica que 50% dos escores encontram-se igual ou abaixo a 63 acertos, sendo a pontuação máxima possível de 120 pontos. Ainda não estão disponíveis dados normativos para a comparação dos escores aqui indicados com os da

população em geral, porém, levando em consideração a faixa a que se destina o TVIP, de 2 anos e 6 meses até 18 anos, considera-se que os alunos que participaram do estudo estejam com o desenvolvimento lexical no domínio receptivo adequados, haja vista a média tenha atingido o ponto médio do instrumento que é 60 pontos.

Tabela 1

Estatísticas descritivas do desempenho dos alunos no Teste de Vocabulário por Imagens – Peabody - TVIP

Estatísticas descritivas das do desempenho no TVIP		Estatística	Erro-padrão
Média		61,95	0,688
Intervalo de confiança de 95% para a média	Limite superior	60,60	
	Limite inferior	63,31	
Mediana		63,00	
Variância		124,021	
Desvio Padrão		11,136	
Mínimo		23	
Máximo		87	
Amplitude total		64	
Amplitude Interquartis		14	
Assimetria		-0,648	0,150
Curtose		0,528	0,300

Verifica-se também, que a curva de frequência se mostra assimétrica negativa (Assimetria = -0,648) e levemente achatada (curtose = 0,528), conforme representada no histograma a seguir (Figura 12). O teste de normalidade de *Kolmogorov-Smirnov* indicou que a distribuição não é compatível com uma curva normal ($K-S [262] = 0,085; p \leq 0,001$). Contudo considera-se que podem ser realizadas análises que necessitam de normalidade, uma vez que a amostra tem número elevado (262 alunos) o que diminui sensivelmente o erro amostral da média.

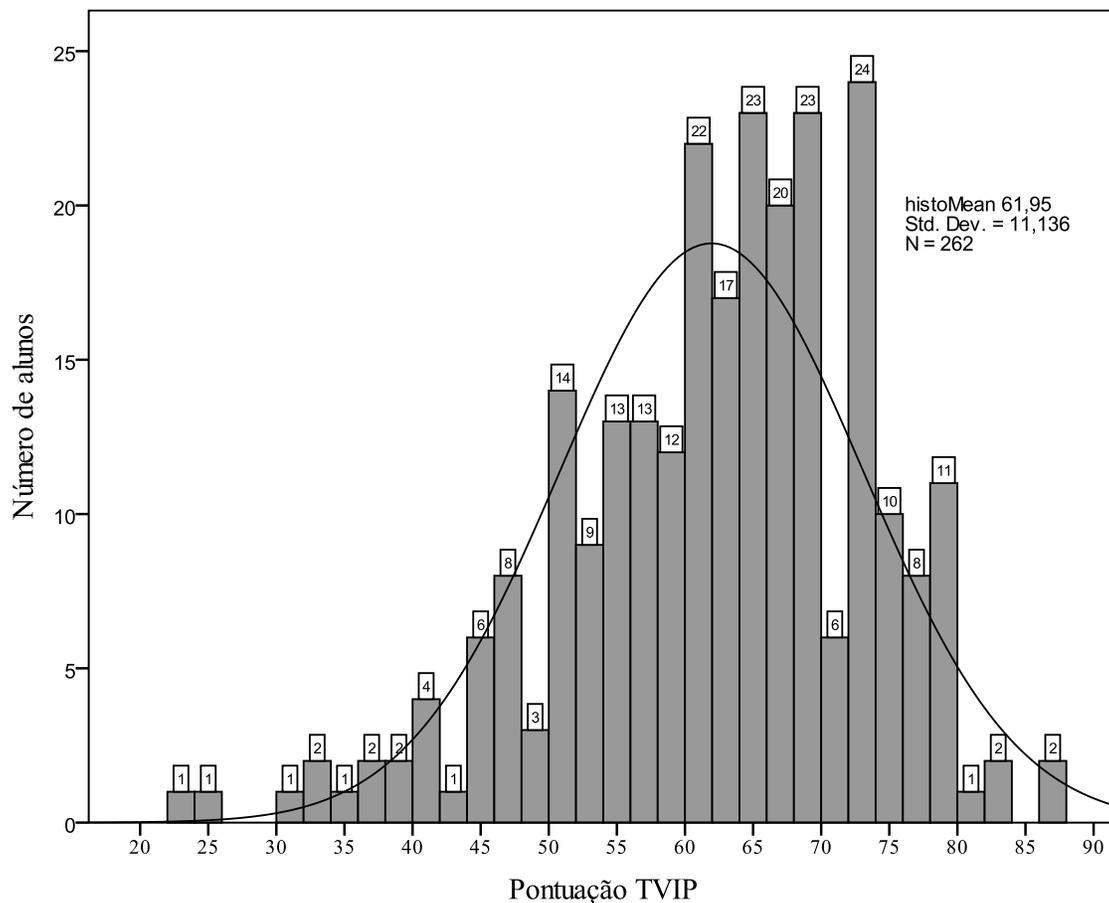


Figura 12. Distribuição de frequência das pontuações dos alunos em vocabulário receptivo.

Com base na posição da mediana (63), pode-se observar que o grupo de estudantes com pontuações entre o percentil 50 e o percentil 25 (63 a 55 acertos) difere mais entre si do que o grupo de estudantes entre o percentil 50 e 75 (63 a 69 acertos). Observa-se no Diagrama de caixa de distribuição (Figura 13) cinco *outliers* indicando os alunos que obtiveram pontuações extremamente baixas, discrepantes do grupo avaliado.

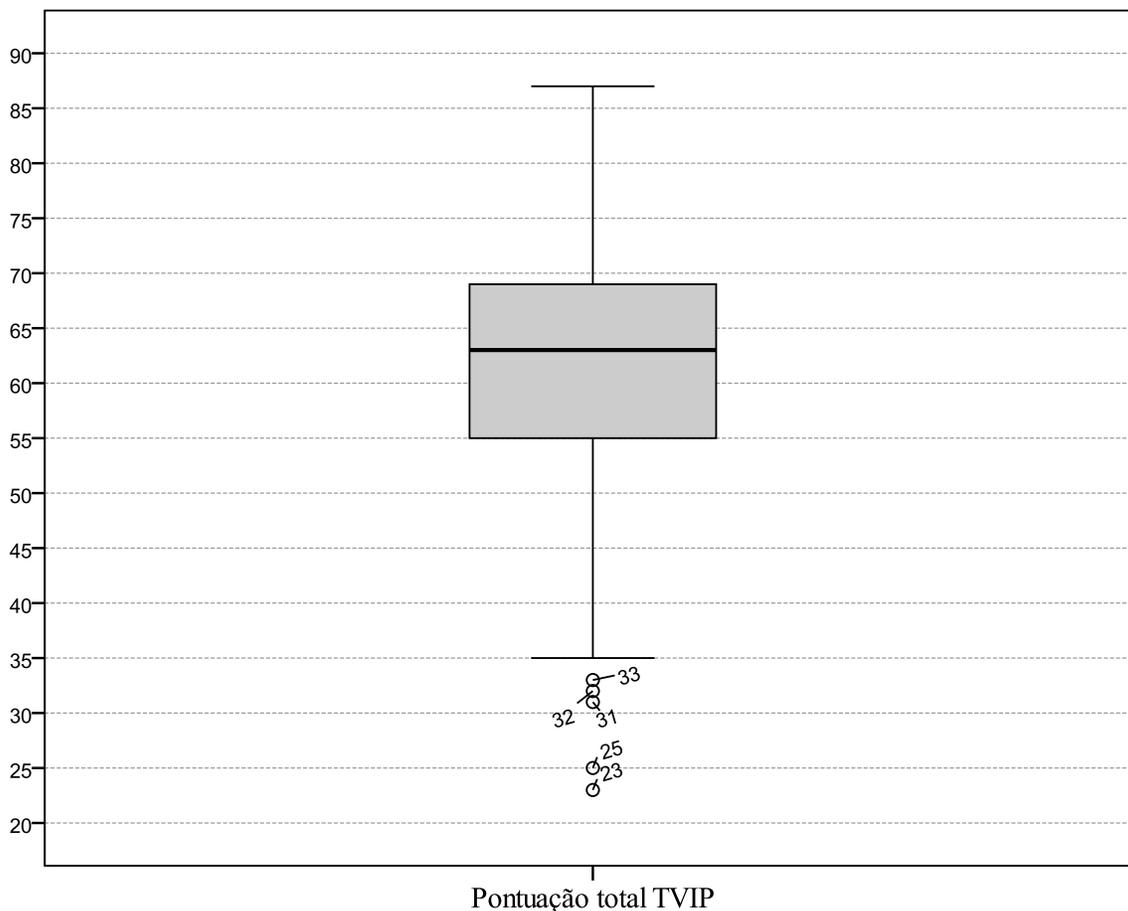


Figura 13. Diagrama de caixa com a distribuição da pontuação em vocabulário receptivo.

Quando observadas as médias no vocabulário receptivo em relação ao ano escolar frequentado (Tabela 2) nota-se que à medida que há progressão da série, também há o aumento das médias. Esta diferença mostra-se estatisticamente significativa ($F[2, 259] = 47,837; p < 0,001$), sendo que, o teste de Tukey confirmou a diferença entre os três anos escolares, isto é, Infantil e 1º ano ($p < 0,001$); Infantil e 2º ano ($p < 0,001$); e entre o 1º e o 2º ano ($p = 0,002$). Deste modo, pode-se afirmar que o instrumento foi sensível ao desenvolvimento acadêmico, ou seja, que as crianças do 2º ano possuem um vocabulário receptivo maior do que as crianças do 1º ano e que estas possuem vocabulário maior do que as do Ensino Infantil, conforme encontrado por Almeida e Joly (2008); Cruz (2011); Istome (2010), Trevisan, Montiel, Dias, e Capovilla (2008).

Tabela 2

Distribuição de médias de vocabulário receptivo e do tempo de execução em função do ano escolar que os alunos estavam matriculados

TVIP	Ano escolar	N	Mínimo	Máximo	Média	Desvio Padrão	$F[2, 259]$	p
Escore	Infantil	85	31	75	54,15	9,476	47,837	<0,001
	1º Ano	96	25	86	63,44	10,248		
	2º Ano	81	23	87	68,38	8,744		
Tempo (segundos)	Infantil	85	343,32	1417,57	604,27	168,595	8,144	<0,001
	1º Ano	96	225,39	1126,66	545,97	137,594		
	2º Ano	81	167,40	910,73	514,71	127,898		
	Total	262	167,40	1417,57	555,22	149,588		

Com relação ao tempo utilizado na execução da prova é possível identificar que à medida que há progressão da série, há a diminuição do tempo gasto para responder. Esta diferença mostrou-se estatisticamente significativa ($F[2, 259] = 8,144$; $p < 0,001$), sendo que, o teste de Tukey indicou que a diferença ocorre entre o último ano da Educação Infantil e os dois anos do Ensino Fundamental, ou seja, infantil e 1º ano ($p = 0,021$); infantil e 2º ano ($p < 0,001$). A diferença do tempo entre o 1º e o 2º ano não foi significativa ($p = 0,331$), podendo ser atribuída ao acaso. Estes dados ratificam os estudos de Istome (2010), que encontrou diferença significativa entre o Infantil e o 1º ano, e o estudo de Cruz (2011) que também não encontrou diferença no tempo de execução do TVIP entre o 1º e o 2º anos do Ensino Fundamental.

Ao observar as diferenças de médias dos alunos em vocabulário receptivo levando-se em consideração a idade (Tabela 3) é possível constatar que, com o aumento da idade

houve o avanço das médias, de modo que, os alunos mais velhos alcançaram escores mais altos que os mais novos. Por meio da ANOVA constatou-se que tais diferenças são estatisticamente significativas ($F[3, 258] = 35,344$; $p < 0,001$). O teste *Post hoc* utilizando *Tukey HSD* revelou que as diferenças encontradas se referiam às idades de 5 e 6 anos ($p < 0,001$), 5 e 7 anos ($p < 0,001$), 5 e 8 anos ($p < 0,001$), 6 e 7 anos ($p < 0,001$) e 6 e 8 anos ($p < 0,001$). Nota-se que, somente não foi encontrada diferença estatisticamente significativa entre 7 e 8 anos ($p = 0,981$), contudo, há que se considerar que o grupo de crianças com 8 anos ($n = 26$) estava com número inferior de participantes, porém próximo de 30 sujeitos. Os resultados obtidos vão ao encontro dos resultados encontrados nas pesquisas de Cruz (2011), Istome (2010), Macedo e cols. (2006) e Trevisan, Montiel, Dias, e Capovilla (2008) que também indicaram efeito significativo da idade sobre o desempenho em vocabulário usando o TVIP computadorizado.

Tabela 3

Distribuição de médias de vocabulário receptivo e do tempo de execução em função da idade dos alunos

TVIP	Idade (anos)	N	Mínimo	Máximo	Média	Desvio Padrão	$F[3, 258]$	p
Escore	5	47	32	73	51,89	9,135	35,344	<0,001
	6	97	25	79	59,63	10,329		
	7	92	23	87	67,70	9,180		
	8	26	49	80	68,50	6,901		
Tempo (segundos)	5	47	343,32	1417,57	638,94	206,713	7,025	<0,001
	6	97	267,06	1126,66	548,81	120,381		
	7	92	167,40	1048,02	521,22	134,440		
	8	26	373,18	888,97	548,10	126,148		

Com relação ao tempo utilizado para realizar o teste, há diminuição do tempo com o avanço da idade, sendo que esta diferença apresenta-se estatisticamente significativa ($F[3, 258] = 7,025; p < 0,001$). O teste *Post hoc* utilizando *Tukey HSD* demonstrou que houve diferenças significativas entre as crianças de 5 e 6 anos ($p = 0,003$), 5 e 7 anos ($p < 0,001$) e 5 e 8 anos ($p = 0,052$), não sendo identificadas diferenças significativas entre as crianças de 6, 7 e 8 anos que estavam com médias muito próximas, ou iguais, no caso de 6 e 8 anos. Cruz (2011), Istome (2010) e Macedo e cols. (2006) encontraram a diminuição do tempo de execução com o avanço da idade sobre o tempo na realização do TVIP.

Os dados encontrados referentes às diferenças de médias no vocabulário receptivo em função do ano escolar e da idade dos alunos, eram esperados, uma vez que, o desenvolvimento psicolinguístico está associado à maturação neurológica que ocorre com o avanço da idade (fator biológico), juntamente com os estímulos ambientais, isto é, influências de ordem social e educacional (Narbona, & Fernández, 2005; Pedroso, & Rotta, 2006). Ao lado disso, segundo Strauss, Sherman e Spreen (2006), pesquisas com *Peabody* indicaram a capacidade do instrumento para diferenciar o desempenho dos alunos em função do desenvolvimento escolar, principalmente entre alunos da Educação Infantil e do Ensino Fundamental, sendo inclusive, considerado como um forte preditor do desenvolvimento da linguagem das crianças.

A análise da diferença de médias dos alunos em função do sexo (Tabela 4) revelou que a média das meninas ($M = 63,68; DP = 9,80$) foi superior a dos meninos ($M = 60,63; DP = 11,93$). O teste *t* de Student comprovou que esta diferença é estatisticamente significativa ($t[258] = 2,269; p = 0,024$). Este achado difere dos resultados das pesquisas de Almeida (2008), Cruz (2011), Istome (2010), Macedo e cols. (2006), que não identificaram diferenças significativas entre o sexo dos alunos. O efeito do sexo não vem sendo reportado ao longo do uso do *Peabody* (Strauss, Sherman & Spreen, 2006) demonstrando ser esta uma

característica da presente amostra. Em relação ao tempo de execução, apesar de os meninos terem sido mais rápidos, esta diferença não se mostrou estatisticamente significativa, podendo, assim, ser atribuída ao acaso, conforme também encontrado nos estudos de Cruz (2011), Istome (2010), Macedo e cols. (2006).

Tabela 4

Distribuição de médias de vocabulário receptivo e do tempo de execução segundo o sexo dos alunos

TVIP	Sexo*	Mínimo	Máximo	Média	Desvio Padrão	gl	t	p
Escore	Feminino	35	87	63,68	9,797	259	2,269	0,024
	Masculino	23	86	60,63	11,931			
Tempo (segundos)	Feminino	225,39	1126,66	556,96	147,770	260	0,165	0,869
	Masculino	167,40	1417,57	553,880	151,463			

*Feminino (n = 114); Masculino (n = 148).

4.2 Desempenho nas Matrizes Progressivas Coloridas de Raven

Assim como realizadas no TVIP, são apresentadas em sequência a análise exploratória, as descritivas e as inferenciais do escore e do tempo de execução no Raven. A análise exploratória apresentada a seguir (Tabela 5) indica que os participantes obtiveram pontuações que variaram entre 6 a 35, sendo que a média foi de 18,16 ($DP = 5,06$). É possível afirmar, com 95% de confiança, que o intervalo da média está entre 17,55 e 18,78. A mediana indica que 50% dos escores encontram-se igual ou abaixo a 18 acertos, ressalta-se que a pontuação máxima possível de se alcançar no teste é de 36 pontos.

Tabela 5

Estatísticas descritivas do desempenho dos alunos no teste Matrizes Progressivas Coloridas de Raven

Estatísticas descritivas das do desempenho no Raven		Estatística	Erro-padrão
Média		18,16	0,313
Intervalo de confiança de 95% para a média	Limite superior	17,55	
	Limite inferior	18,78	
Mediana		18	
Variância		25,601	
Desvio Padrão		5,060	
Mínimo		6	
Máximo		35	
Amplitude total		29	
Amplitude Interquartis		6	
Assimetria		0,446	0,150
Curtose		0,446	0,300

Verifica-se também que a curva de frequência se mostra assimétrica positiva (0,446) e levemente achatada (curtose = 0,446), conforme representada no histograma a seguir (Figura 14). O teste de normalidade de *Kolmogorov-Smirnov* indicou que a distribuição não é compatível com uma curva normal ($K-S [262] = 0,070$; $p = 0,003$), assim como observado no TVIP. Do mesmo modo que no teste anterior, considera-se que podem ser realizadas análises que necessitam de normalidade, uma vez que a amostra tem número elevado de participantes (262 alunos) o que diminui sensivelmente o erro amostral da média.

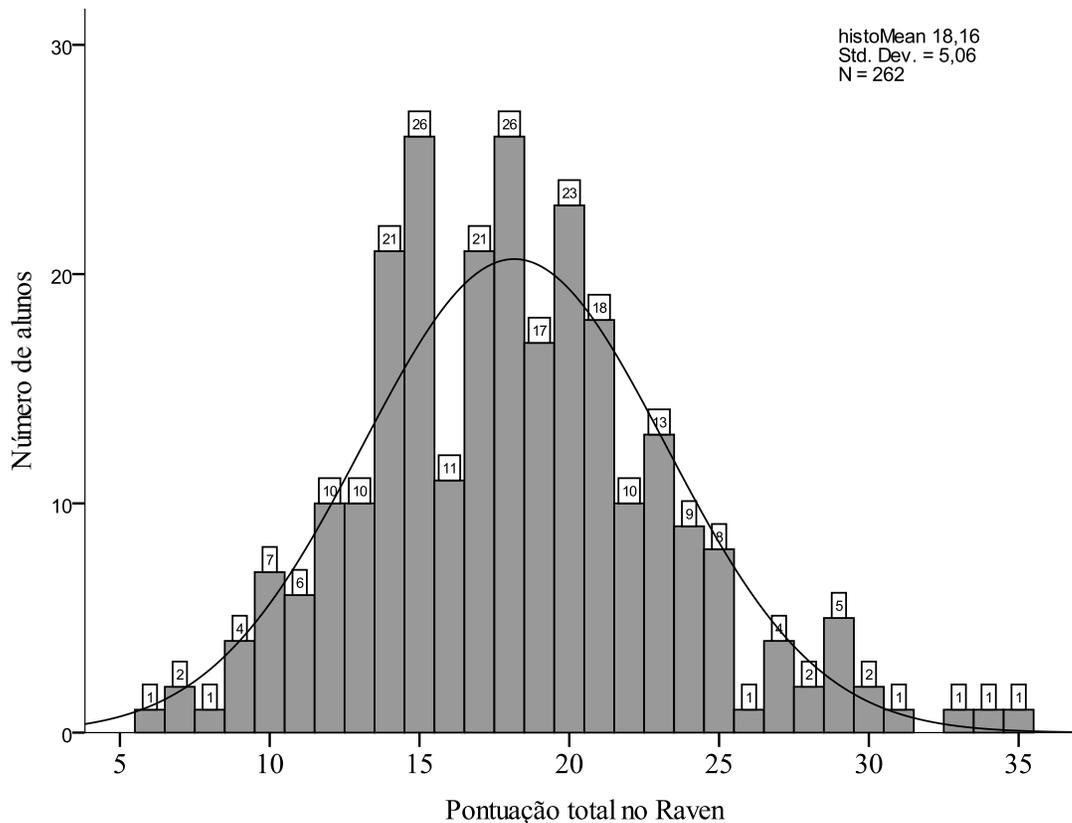


Figura 14. Distribuição de frequência das pontuações dos alunos em inteligência.

Quando observadas as pontuações no Raven para cada ano escolar avaliado, é possível visualizar no Diagrama de caixa apresentado a seguir (Figura 15) o aumento na pontuação com a escolarização. Nota-se que 50% dos alunos do Ensino Infantil obtiveram pontuações iguais ou menores a 15 pontos; metade dos alunos do 1º ano obtiveram pontuações iguais ou menores a 18 pontos; e dos alunos do 2º ano iguais ou menores que 20, conforme indicam as medianas. Ressalta-se a existência de alunos *outliers*, ou seja, com pontuações mais altas do que o padrão do grupo.

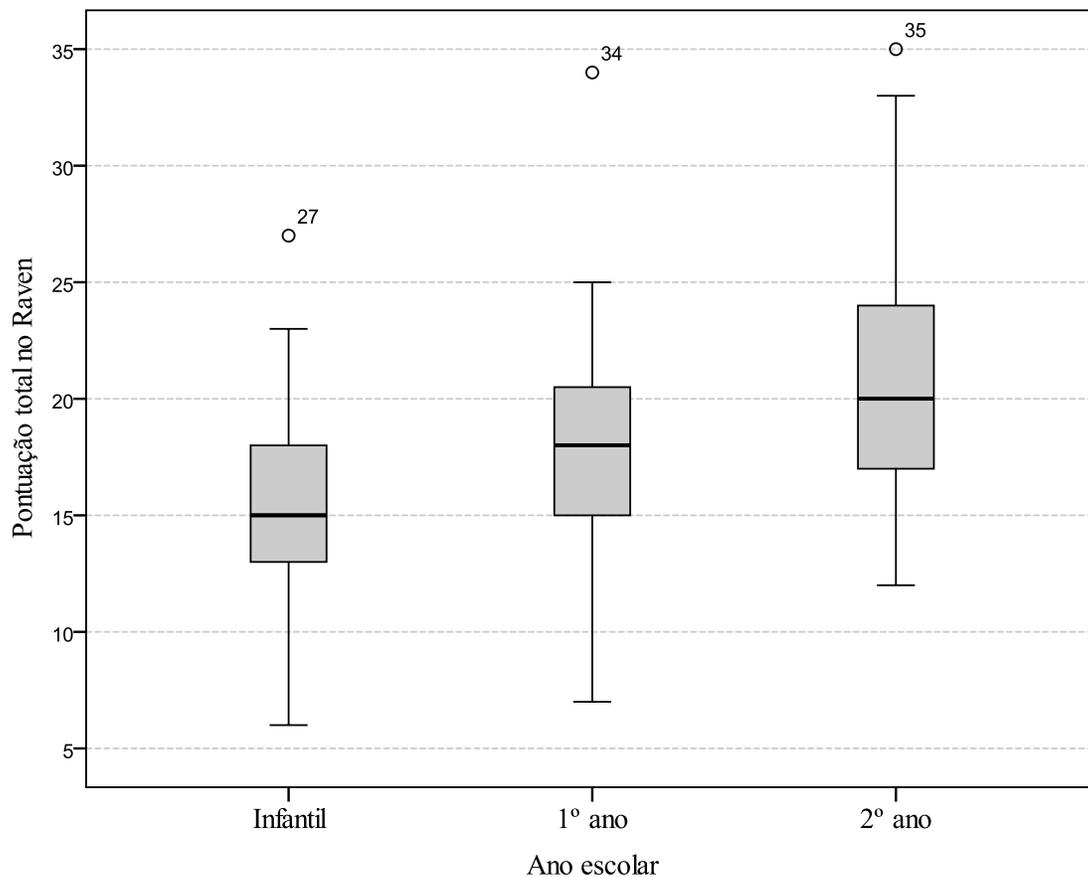


Figura 15. Diagrama de Caixa com a distribuição de frequência das pontuações dos alunos em inteligência por ano escolar.

Ao analisar as médias dos alunos em relação ao ano escolar que se encontravam matriculados, o teste de homogeneidade de *Levene* indicou que as variâncias observadas são homogêneas, tanto para escore de inteligência ($F[2, 259] = 2,676; p = 0,071$) quanto para o tempo de execução ($F[2, 257] = 0,878; p = 0,417$). Da observação das médias dos alunos em função do ano escolar (Tabela 6), constata-se que há um aumento das médias com a progressão escolar. A ANOVA indicou que tais diferenças são estatisticamente significativas, o teste *Post Hoc* usando o *Tukey HSD* identificou que houve diferença de médias dos escores de todos os anos, ou seja, entre o Infantil e o 1º ano ($p = 0,002$); o Infantil e 2º ano; ($p \leq 0,001$) e entre o 1º e 2º ano ($p \leq 0,001$). Esta diferenciação já era prevista, conforme estudos apresentados no manual do teste (Angelini e cols., 1999).

Tabela 6

Distribuição de médias de inteligência e do tempo de execução em função do ano escolar que os alunos estavam matriculados

Raven	Ano escolar	N	Mínimo	Máximo	Média	Desvio Padrão	$F[2, 259]$	p
Escore	Infantil	85	6	27	15,58	4,098	30,951	$\leq 0,001$
	1º Ano	96	7	34	17,95	4,421		
	2º Ano	81	12	35	21,14	5,154		
	Total	262	6	35	18,16	5,060		
Tempo (minutos)	Infantil	85	4	10	5,95	1,575	8,342	$\leq 0,001$
	1º Ano	96	4	13	6,99	1,759		
	2º Ano	81	4	19	6,94	2,249		
	Total	262	4	19	6,64	1,925		

Em relação ao tempo utilizado para a execução do teste, este foi aferido pelo aplicador do teste utilizando-se o relógio. Por esta razão o tempo foi contado em minutos, diferentemente dos outros testes utilizados que eram informatizados. Também na Tabela 6, é possível verificar que, assim como houve aumento das médias dos escores de inteligência com avanço da série, houve também o aumento do tempo gasto para responder ao teste. Novamente a ANOVA mostrou que esta diferença é estatisticamente significativa, sendo que, o teste *Post Hoc* utilizando o *Tukey HSD* indicou que houve diferença entre o Infantil e o 1º ano ($p \leq 0,001$) e o Infantil e o 2º ano ($p = 0,002$), não havendo diferença entre o 1º ano e o 2º ano ($p = 0,982$).

De modo a tornar mais claros os dados obtidos nas pontuações, a seguir será exposta a classificação dos resultados, ou seja, a interpretação do nível de inteligência da criança com base na normatização brasileira para alunos de escolas públicas, conforme o manual do teste (Angelini e cols., 1999). Na Figura 16 é possível verificar que há predominância de

alunos “Intelectualmente médios” no Ensino Infantil (36,5%) e no 2º ano (44,4%), seguidos por “Definidamente acima da média na capacidade intelectual” e “Intelectualmente Superior”. Já no 1º ano prevalecem alunos com a classificação “Definidamente acima da média na capacidade intelectual” (42,7%). Ressalta-se que há mais alunos situados nos níveis acima da média do que nos níveis abaixo em todos os anos escolares, indicando ser uma amostra com bom nível cognitivo, quando comparada com a população de referência do teste.

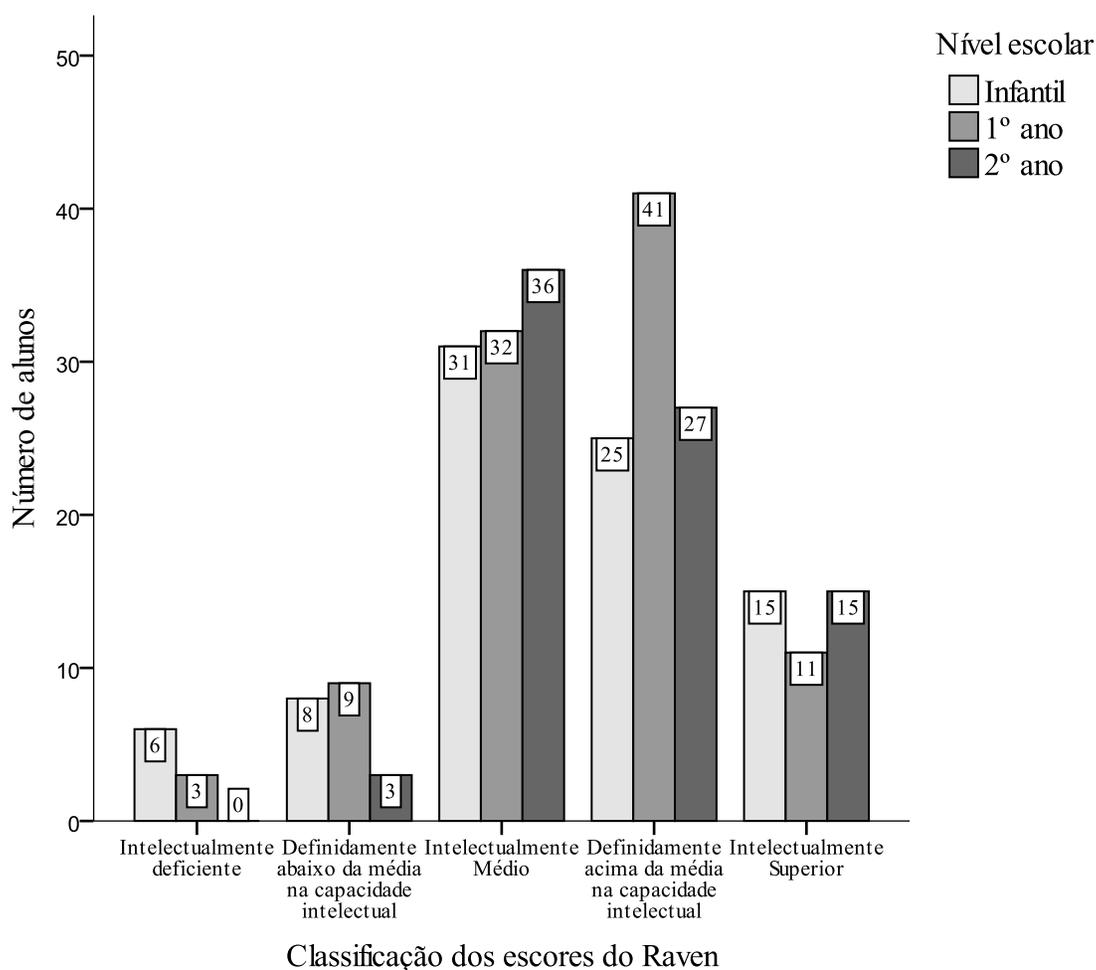


Figura 16. Distribuição de frequência da classificação dos alunos segundo o nível de inteligência, por ano escolar.

De modo a verificar a distribuição das médias de inteligência em função da idade dos alunos, a Tabela 7 torna claro que há o aumento das médias com o aumento da idade,

bem como o aumento da variabilidade, observada nos desvios padrão. A ANOVA revelou ser estatisticamente significativa esta diferença e o teste *Post hoc* utilizando-se *Tukey HSD* indicou que as idades que diferem são entre 5 e 7 ; 5 e 8; e 6 e 7 anos com índices de significância menores do que 0,001. Não diferiram pelo teste *Post hoc* as idades 5 e 6 ($p = 0,401$) e 8 e 7 ($p = 0,352$).

Tabela 7

Distribuição de médias de inteligência e do tempo de execução no Raven em função da idade dos alunos

Raven	Idade (anos)	N	Mínimo	Máximo	Média	Desvio Padrão	$F[3, 258]$	p
Escore	5	47	6	23	15,36	4,002	21,208	$\leq 0,001$
	6	97	7	27	16,63	4,421		
	7	92	12	35	20,17	4,771		
	8	26	12	33	21,85	5,198		
Tempo (minutos)	5	47	4	10	5,87	1,600	3,266	0,022
	6	97	4	13	6,71	1,820		
	7	92	4	14	6,84	1,815		
	8	26	5	19	7,04	2,793		

O esperado, conforme Angelini e cols. (1999) é que houvesse a diferenciação estatisticamente significativa entre todas as idades. Há que se considerar que no manual do teste a idade é referida por faixas etárias, sendo que cada faixa tem amplitude de 6 meses (ex. a faixa de idade referente a 5 anos, engloba crianças com 4 anos e 9 meses a 5 anos, 2 meses e 29 dias; sendo a próxima faixa etária referente a 5 $\frac{1}{2}$ anos, e assim por diante). Como no presente estudo foram analisadas as idades em anos, este fato pode ter sido responsável pela não diferenciação, uma vez que, os alunos de 5 anos estavam por fazer 6 e os de 8 teriam acabado de fazer aniversário. Deste modo, entende-se que o ano escolar foi

melhor diferenciador do desempenho dos alunos no Raven, já que ele diferenciou os três anos escolares, sendo que as idades de 5 e 6 anos compreendem os alunos do Infantil, e as idades 7 e 8 compreendem os alunos do 2º ano.

Assim como encontrado na observação dos resultados por ano escolar frequentado, olhando-se para as diferenças de médias de tempo em função da idade do aluno também foi constatado que estas se apresentam estatisticamente significativas. Entretanto estas diferenças se deram somente entre as idades de 5 e 7 anos ($p = 0,027$), sendo que para os demais anos estas diferenças podem ser atribuídas ao acaso: 5 e 6 ($p = 0,066$); 5 e 8 ($p = 0,061$); 6 e 7 ($p = 0,970$); 6 e 8 ($p = 0,864$); e 7 e 8 ($p = 0,963$).

A análise da diferença de médias dos alunos em função do sexo revelou que a média dos escores das meninas ($M = 18,41$; $DP = 4,63$) foi superior a dos meninos ($M = 17,97$; $DP = 5,37$). Contudo, esta diferença não foi estatisticamente significativa ($t [260] = 0,696$; $p = 0,487$), podendo ser atribuída ao acaso (Tabela 8).

Tabela 8

Distribuição de médias de inteligência e do tempo de execução do Raven segundo o sexo dos alunos

Raven	Sexo*	Mínimo	Máximo	Média	Desvio Padrão	gl	t	p
Escore	Feminino	7	33	18,41	4,633	260	0,696	0,487
	Masculino	6	35	17,97	5,373			
Tempo (minutos)	Feminino	4	14	6,64	1,808	258	-0,009	0,992
	Masculino	4	19	6,64	2,017			

*Feminino (n = 114); Masculino (n = 148).

Na pesquisa normativa do instrumento também não houve diferença significativa por sexo, a não ser para as idades de 11 anos, que não é o caso da presente pesquisa (Angelini e cols., 1999). Em relação ao tempo de execução, não foi encontrada diferença na média, ambos obtiveram média de 6,64 minutos, sendo que houve maior variabilidade de médias de tempo para os meninos ($DP = 2,017$) do que para as meninas ($DP = 1,808$).

Foi verificada a associação do escore com o tempo de execução no Raven por meio do teste de correlação de *Pearson* e os resultados indicaram fraca correlação ($r = 0,230$), porém muito significativamente diferente de zero ($p \leq 0,001$), indicando que há uma tendência ao aumento de pontos com o aumento do tempo. Conforme observado durante as aplicações, em geral, os alunos que demoraram mais tempo buscando a lógica para a resolução do problema, obtiveram melhores desempenho.

Lembrando que a educação, é um processo de “*extrair novas informações daquilo que é percebido ou conhecido*” (Angelini e cols., 1992, p. 5), o processo pelo qual se dá esta educação, tem como a primeira ação a percepção do objeto global para que se possa entender o problema solicitado e passe-se, então, a investigar as relações, monitorando as tentativas de solução consistente com todas as informações percebidas. Para a resolução do Raven são ainda requeridas as representações simbólicas de linhas, círculos, quadrados e triângulos. Deste modo, a capacidade realizar estas provas envolve a experiência e a aprendizagem cultural (Angelini e cols., 1992). Isto explica o aumento da pontuação e do tempo de execução em função da progressão do ano escolar e com o avanço da idade, pois com o desenvolvimento há também a ampliação desta capacidade analítica e o aumento das hipóteses para a solução dos problemas, que demandam mais tempo e promovem maiores chances de acerto.

4.3 Desempenho na Bateria Informatizada de Linguagem Oral, BILOv3

Assim como realizados nos dois instrumentos descritos anteriormente, procedeu-se, inicialmente a análise exploratória do total da BILOv3 . O teste de normalidade indicou que a distribuição dos escores é diferente da curva normal (K-S [262] = 0,113; $p < 0,001$). Assim como nos outros testes, considerou-se que, devido ao tamanho considerável da amostra as posteriores análises não seriam prejudicadas. Na Tabela 9 constam os dados referentes as estatísticas descritivas da BILOv3.

Tabela 9

Estatísticas da análise exploratória do desempenho dos alunos na BILOv3

Estatísticas descritivas das do desempenho da BILOv3	Estatística	Erro-padrão
Média	87,48	1,039
Intervalo de confiança de 95% para a média	85,44 89,53	
Mediana	84,00	
Variância	282,772	
Desvio Padrão	16,816	
Mínimo	48	
Máximo	118	
Amplitude total	70	
Amplitude Interquartis	27	
Assimetria	0,122	0,150
Curtose	- 0,826	0,300

Dos resultados disponíveis na Tabela 9, nota-se que a média na pontuação total da BILOv3, que é a somatória das notas de todos as provas da bateria, foi igual a 87,48 ($DP = 16,82$), sendo que há 95% de chance desta média estar entre 85,44 e 89,53. Os escores variaram de 48 a 118 pontos, sendo que a moda foi de 83 pontos, e a mediana indicou que 50% dos alunos obtiveram médias iguais ou menor que 84 pontos. A curva

formada pela distribuição dos escores ficou próxima da simetria da curva normal, sendo o coeficiente de assimetria igual a 0,12, já a o coeficiente da curtose (- 0,826) indicou que a curva está mais achatada do que a curva normal. Tais dados são melhor visualizados na Figura 17.

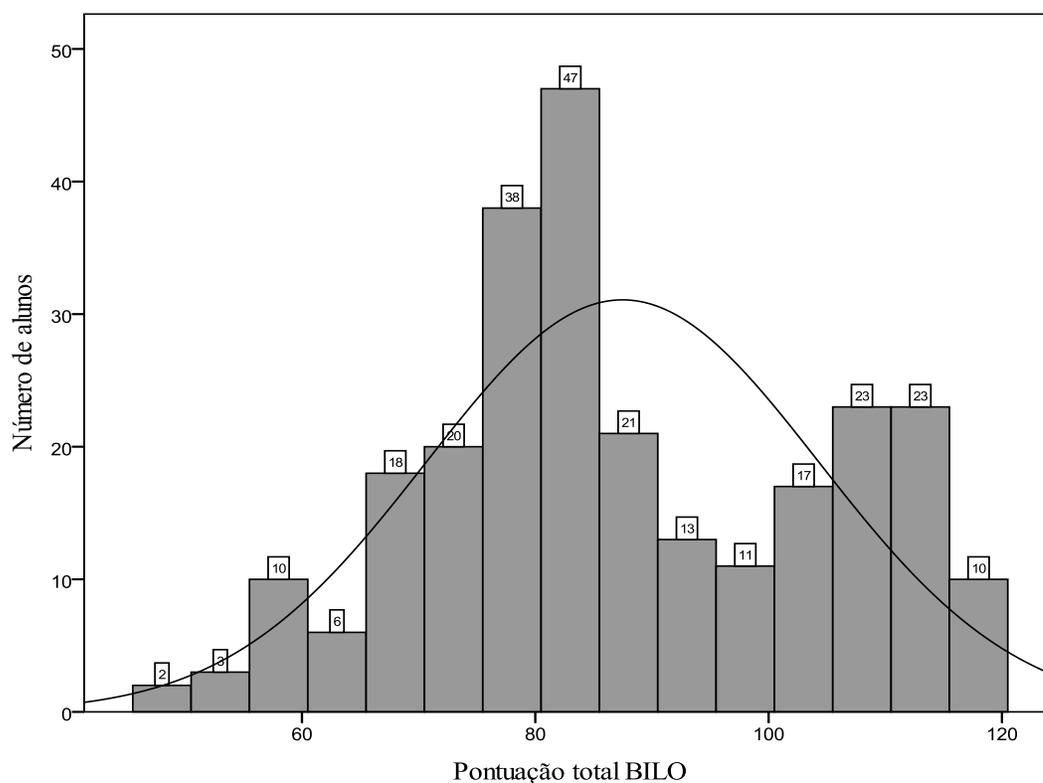


Figura 17. Distribuição da pontuação em compreensão da linguagem oral.

Como forma de verificar como se distribuam os escores dos alunos em cada prova da BILOv3, foi elaborado um Diagrama de caixa, conforme Figura 18. Da observação das caixas referentes às provas da BILOv3 constata-se que as provas CM, CH e IH apresentaram efeito teto, ou seja, os escores apresentaram-se demasiadamente altos. Na prova CM, IH e CH a moda, ou seja, o resultado que apareceu com maior frequência foi 20 pontos, que se refere à nota máxima de cada prova. No caso da CH a mediana também foi 20 pontos, indicando que metade dos alunos atingiram resultados iguais ou inferiores ao máximo possível para a prova. As amplitudes também chamam a atenção, enquanto a SL e

assimetria negativa das provas, com destaque para as provas CH (-2,373), CM (-1,767), CF (-1,825) e IH (-1,152) confirmando os dados já observados nos diagramas de caixa. Por meio da observação da Curtose destacam-se as provas CM (3,145), CF (4,184) e CH (6,395) com índices muito elevados.

Tabela 10

Estatísticas descritivas do desempenho nas provas da BILOv3

Provas	Média	Desvio Padrão	Assimetria		Curtose	
			Estatística	Desvio Padrão	Estatística	Desvio padrão
CM	17,84	2,742	-1,767	0,150	3,145	0,300
SL	7,79	5,979	0,749	0,150	-1,027	0,300
OLV	8,53	5,475	0,720	0,150	-0,777	0,300
IH	17,07	3,344	-1,152	0,150	0,424	0,300
CF	17,47	2,517	-1,825	0,150	4,184	0,300
CH	18,77	2,107	-2,373	0,150	6,395	0,300
BILO	87,48	16,816	0,122	0,150	-0,826	0,300

Tendo em vista que avaliação e o diagnóstico em linguagem oral devam ser baseados em critérios que auxiliem os profissionais a caracterizarem o desempenho e identificar os aspectos deficitários do avaliando (Chevrie-Muller & Narbona, 2005), a identificação das áreas específicas avaliadas pelas provas CM, CF, CH podem não estar sendo efetivas para as idades e anos escolares abarcados pelo Ensino Fundamental. Ressalta-se que as referidas provas com defasagem são de fundamental importância para a avaliação da compreensão da linguagem oral.

A prova Compreensão Morfossintática, refere-se à unidade de tratamento “sintaxe” e à “morfologia”, que são responsáveis pelo processo de combinação de regras gramaticais que regulam a união de sílabas em palavras e destas em frases, e deste ponto, promovem a

identificação lexical, que corresponde à significação da palavra que foi ouvida, a partir do acionamento do conhecimento prévio que o ouvinte possua. Já as provas CF e CH, visam ascender ao nível de processamento cognitivo que envolve a “Semântica”, que engloba as regras gramaticais para a composição de frases, orações e a “Pragmática”, que se refere à contextualização do que foi ouvido, e que levam à efetiva compreensão da linguagem oral (Chevrie-Muller & Narbona, 2005).

Como forma de verificar como as médias das provas e do total da *BILOv3* se apresentaram ao longo dos 5 estudos realizados com a terceira versão desta bateria, na Tabela 11 foram organizadas as médias e desvios-padrão dos resultados obtidos anteriormente, juntamente com os do presente estudo. Ressalta-se que, conforme legenda anexada à tabela, as pesquisas compreendem amostras de diferentes idades (de 4 a 13 anos) e de anos escolares que vão desde o último ano da Educação Infantil até o quinto ano do Ensino Fundamental, de escolas públicas e privadas. Na última coluna estão os dados encontrados na presente pesquisa, sob denominação “Atual”. Destaque é dado ao fato de que as pesquisas relatadas foram realizadas no estado de São Paulo, na capital (Istome, 2010) e no interior (Istome, 2010; Joly e cols, 2010; Cruz, 2012 e a atual pesquisa), no Rio Grande do Norte (Freitas, 2011) e o Amapá (Soares, 2011), compreendendo, assim, amostras das regiões Sudeste, Nordeste e Norte do país.

Apesar das diferenças observadas entre as características das amostras das cinco pesquisas, há pouca diferença entre os resultados dos estudos. É possível observar na Tabela 11 que nas provas CM, IH, CF e CH as médias dos alunos diferem em poucos pontos, em geral ficam entre 17 e 19 pontos. Já nas provas SL e OLV, os valores variam em relação à idade média de cada amostra, alcançando a maior média (16,27) na prova SL, para alunos de escola privada com idade média de 8,06 anos (Joly e cols, 2010), e a menor

($M = 7,69$) em alunos com idade média de 5,37 anos ($DP = 0,59$) no estudo de Istome (2010).

Tabela 11

Estatísticas descritivas do desempenho nas provas da BILOv3 nos cinco estudos anteriores e na presente pesquisa

Provas		Istome (2010) ¹	Joly e cols. (2010) ²	Freitas (2011) ³	Soares (2011) ⁴	Cruz (2012) ⁵	Atual ⁶
CM	<i>M</i>	18,20	19,55	19,11	17,76	18,95	17,84
	<i>DP</i>	2,60	1,08	3,08	3,20	2,00	2,742
SL	<i>M</i>	7,69	16,27	13,27	8,51	13,74	7,79
	<i>DP</i>	5,54	4,29	6,34	6,44	5,80	5,98
OLV	<i>M</i>	8,38	16,79	13,40	9,40	13,86	8,53
	<i>DP</i>	5,25	4,19	6,37	5,90	5,69	5,47
IH	<i>M</i>	16,87	17,50	18,32	17,54	18,75	17,07
	<i>DP</i>	3,40	3,19	3,97	3,07	2,10	3,34
CF	<i>M</i>	17,21	19,16	18,13	17,51	18,88	17,47
	<i>DP</i>	3,03	1,22	4,43	2,27	1,56	2,52
CH	<i>M</i>	18,44	18,27	18,17	18,94	19,55	18,77
	<i>DP</i>	2,14	3,19	4,92	1,98	1,07	2,107
Total	<i>M</i>	86,78	NC	100,39	89,66	103,72	87,48
	<i>DP</i>	15,89	NC	23,44	17,44	14,17	16,82

¹ n = 157 alunos do último ano da Ed. Infantil e 1º ano de três escolas privadas, com idades entre 4 e 7 anos ($M = 5,37$ e $DP = 0,59$)

² n = 154 alunos do 2º ao 5º ano de uma escola privada, com idades entre 6 e 10 anos ($M = 8,06$; $DP = 1,13$)

³ n = 101 alunos do 1º ao 3º ano de uma escola pública, idades entre 5 e 8 anos ($M = 6,81$; $DP = 0,89$)

⁴ n = 290 alunos do 1º ao 4º ano de uma escola pública, idades entre 5 e 9 anos ($M = 7,12$; $DP = 1,23$)

⁵ n = 325 alunos do 1º a 5º ano de uma escola pública, idades entre 6 e 13 anos ($M = 9,03$; $DP = 1,535$)

⁶ n = 262 alunos do último ano da Ed. Infantil e 1º ano de escolas públicas, com idades entre 5 e 8 anos ($M = 6,37$ e $DP = 0,89$)

NC = não consta

Na sequência, foi realizada uma avaliação das médias dos escores por prova e total da BILOv3 levando em conta o ano escolar em que os alunos se encontravam. Os resultados constam na Tabela 12.

Tabela 12

Estatísticas descritivas dos escores por prova e total da BILOv3 em função do ano escolar.

Escore	Ano escolar	N	Mínimo	Máximo	Média	Desvio Padrão	$F[2, 259]$	p
CM	Infantil	85	5	20	16,29	3,548	27,613	$\leq 0,001$
	1º Ano	96	12	20	18,14	2,106		
	2º Ano	81	13	20	19,12	1,317		
	Total	262	5	20	17,84	2,742		
SL	Infantil	85	1	19	4,16	3,330	50,272	$\leq 0,001$
	1º Ano	96	1	19	7,41	5,463		
	2º Ano	81	0	20	12,06	6,077		
	Total	262	0	20	7,79	5,979		
OLV	Infantil	85	1	18	5,75	3,027	35,487	$\leq 0,001$
	1º Ano	96	0	20	8,03	5,250		
	2º Ano	81	0	20	12,05	5,865		
	Total	262	0	20	8,53	5,475		
IH	Infantil	85	7	20	15,05	3,501	30,087	$\leq 0,001$
	1º Ano	96	6	20	17,63	3,116		
	2º Ano	81	9	20	18,53	2,281		
	Total	262	6	20	17,07	3,344		
CF	Infantil	85	6	20	16,27	3,118	21,123	$\leq 0,001$
	1º Ano	96	8	20	17,55	2,243		
	2º Ano	81	16	20	18,63	1,259		
	Total	262	6	20	17,47	2,517		
CH	Infantil	85	8	20	17,74	2,842	16,927	$\leq 0,001$
	1º Ano	96	11	20	19,22	1,537		
	2º Ano	81	15	20	19,32	1,233		
	Total	262	8	20	18,77	2,107		
Total BILO	Infantil	85	49	117	75,27	12,681	65,623	$\leq 0,001$
	1º Ano	96	48	118	87,97	14,233		
	2º Ano	81	67	118	99,72	14,243		
	Total	262	48	118	87,48	16,816		

Os resultados revelam que há aumento progressivo das médias do Infantil ao 2º Ano em todas as provas e no valor total. Observam-se também valores de desvios padrão

distribuídos de modo crescente entre os anos escolares, a saber, Infantil, 1º e 2º anos nas provas CM, SL, OLV e de modo decrescente nas provas IH, CF e CH, revelando que nas três primeiras provas há maior variabilidade de médias para o Infantil do que para o 2º ano e que nas últimas três provas esta situação se inverte.

Novamente, podem ser visualizadas as menores médias e os maiores desvios padrão nas provas SL e OLV, no entanto, acrescenta-se a informação de que a diferença entre os anos escolares é bem delimitada principalmente do 1º para o 2º anos, que diferem em mais de quatro pontos entre uma série e outra. As demais provas apresentam pequenas diferenças entre as médias dos diferentes anos escolares.

Para verificar o nível de significância destas diferenças foi realizada a análise de variância por meio da ANOVA, que evidenciou que houve diferença estatisticamente significativa para todas as provas e total da BILOv3, indicando que há diferenciação no desempenho em linguagem oral em função da escolarização. Como forma de averiguar em que anos escolares as diferenças se encontravam foi realizado o teste *post hoc* utilizando o *Tukey HSD*. Desta forma, foi possível identificar que houve a formação de três grupos que explicam a variância dos escores entre as provas CM ($p \leq 0,025$), SL ($p \leq 0,001$), OLV ($p \leq 0,005$), CF ($p \leq 0,007$) e para o total da BILOv3 ($p \leq 0,001$), indicando que estas provas diferenciaram os alunos do Infantil, do 1º ano e o 2º ano. Nas provas IH ($p \leq 0,001$) e CH ($p \leq 0,001$) houve a formação de apenas dois grupos, sendo um grupo para o Infantil e outro para o 1º e 2º anos do Fundamental, evidenciando a diferença entre estes dois níveis de ensino. As diferenças entre 1º e 2º anos não foram estatisticamente significativas nas provas IH ($p \leq 0,118$) e CH ($p \leq 0,938$).

As diferenças no desempenho da BILO em função do ano escolar foram observadas nos estudos com a primeira (Almeida & Joly, 2008; Issa, 2008; Joly & Piovezan, 2012) e com a segunda versão (Joly, Piovezan, Soares, Lopes & Martins, 2009; Joly, Martins,

Lopes & Lemos, 2009), e mantiveram-se em todos os estudos realizados com a terceira versão (Istome, 2010; Joly e cols., 2010; Cruz, 2011; Freitas, 2011; Soares, 2011), sempre indicando que com o avanço da escolarização há o aumento do desempenho em compreensão da linguagem oral. Estes resultados atribuem à BILO validade de critério por escolaridade (AERA, APA & NMCE, 1999; Anastasi & Urbina, 2000).

Em relação à possível diferenciação do desempenho na BILOv3 em função da variável idade, também se procedeu a verificação das diferenças de médias por meio da ANOVA, que apontou serem estatisticamente significativas. A Tabela 13 expressa as estatísticas descritivas e os resultados da ANOVA com os valores de F e p para os escores da BILOv3.

Há que se constar que a análise preliminar de igualdade das variâncias entre os grupos, por meio do teste de Levene mostrou na pontuação total da BILOv3 que as variâncias dos erros dos grupos são iguais ($p = 0,160$), ou seja há homogeneidade. O teste de *post hoc* utilizando o Tukey HSD identificou que na prova CM ($p \leq 0,05$) e na pontuação total da BILOv3 ($p \leq 0,05$) houve a formação de três grupos, quais sejam, alunos com 5 anos; alunos com 6 anos e, conjuntamente, alunos de 7 e 8 anos, indicando a diferenciação entre estes três grupos, com pontuações crescentes. Na prova SL ($p \leq 0,05$), também houve a formação de três grupos, sendo que, no primeiro grupo ficaram os alunos de 5 e 6 anos; no segundo alunos com 7 anos e no terceiro alunos com 8 anos. Nas provas OLV ($p \leq 0,001$), IH ($p \leq 0,05$) e CP ($p \leq 0,002$), dois grupos se formaram, diferenciando alunos de 5 e 6 anos dos alunos com 7 e 8 anos. Por fim, na prova CH houve a diferenciação dos alunos de 5 anos para os demais, com 6, 7 e 8 anos.

Estes resultados revelam evidências de validade da BILO para o critério relativo ao desenvolvimento, uma vez que o instrumento diferencia alunos mais novos dos alunos mais velhos (AERA, APA & NMCE, 1999; Anastasi & Urbina, 2000), corroborando os cinco

estudos com a BILOv3 (Istome, 2010; Joly e cols., 2010; Cruz, 2011; Freitas, 2011; Soares, 2011). A capacidade de diferenciar o desempenho dos alunos ao longo do desenvolvimento também tem sido constante desde os estudos com as versões anteriores (Almeida & Joly (2008); Issa (2008); e Joly e Piovezan (2012); Joly, Reppold, & Dias 2009; Joly, Martins, Lopes, & Lemos, 2009).

Tabela 13

Estatísticas descritivas dos escores por prova e total da BILOv3 em função da idade.

Escore	Idade (anos)	N	Mínimo	Máximo	Média	Desvio Padrão	$F[3, 258]$	p
CM	5	47	5	20	15,89	3,737	16,660	$\leq 0,001$
	6	97	9	20	17,52	2,762		
	7	92	13	20	18,87	1,570		
	8	26	16	20	18,96	1,341		
SL	5	47	1	19	3,96	3,196	28,800	$\leq 0,001$
	6	97	1	19	5,99	5,086		
	7	92	0	20	10,01	5,972		
	8	26	2	20	13,62	5,629		
OLV	5	47	1	18	5,60	2,826	18,873	$\leq 0,001$
	6	97	0	19	7,07	4,537		
	7	92	0	20	10,35	6,050		
	8	26	0	20	12,88	5,443		
IH	5	47	8	20	15,09	3,611	14,158	$\leq 0,001$
	6	97	6	20	16,47	3,606		
	7	92	9	20	18,32	2,233		
	8	26	9	20	18,46	2,860		
CF	5	47	5	20	15,89	3,737	12,176	$\leq 0,001$
	6	97	7	20	16,71	2,922		
	7	92	15	20	18,37	1,404		
	8	26	16	20	18,69	1,379		
CH	5	47	8	20	17,70	2,773	6,584	$\leq 0,001$
	6	97	9	20	18,70	2,341		
	7	92	14	20	19,29	1,245		
	8	26	16	20	19,12	1,479		
Total BILO	5	47	52	117	74,83	12,478	33,601	$\leq 0,001$
	6	97	48	115	82,46	15,188		
	7	92	67	118	95,21	14,150		
	8	26	67	118	101,73	14,888		

Quando observadas as diferenças de médias no desempenho em linguagem oral em função do sexo do aluno, observou-se que as meninas obtiveram médias mais altas que os meninos em todas as provas e no total da BILOv3. Entretanto, o teste *t* de *Student* indicou que as diferenças foram estatisticamente significativas somente nas provas SL, OLV, IH e no total da BILOv3, conforme demonstrado na Tabela 14. Ao investigar em quais anos escolares esta diferença se mostrou descobriu-se que as diferenças das provas SL, OLV e Total se deram somente entre os alunos do 2º ano ($t_{SL}[72, 550] = 2,442$; $p = 0,018$; $t_{OLV}[79] = 2,006$; $p = 0,048$; $t_{Total} [79] = 2,418$; $p = 0,018$) e a diferença na provas IH entre os alunos do Infantil.

Tabela 14

Estatísticas descritivas dos escores por prova e total da BILOv3 segundo o sexo do aluno

Escore	Sexo*	Média	Desvio Padrão	gl	t	p
CM	Feminino	18,11	2,787	260	1,404	0,161
	Masculino	17,64	2,698			
SL	Feminino	9,38	6,320	222,332	3,784	≤0,001
	Masculino	6,57	5,416			
OLV	Feminino	9,75	5,675	230,314	3,157	0,002
	Masculino	7,60	5,143			
IH	Feminino	17,90	2,733	259,887	3,761	≤0,001
	Masculino	16,43	3,627			
CF	Feminino	17,76	2,497	260	1,663	0,097
	Masculino	17,24	2,517			
CH	Feminino	18,96	1,848	259,317	1,345	0,180
	Masculino	18,62	2,281			
Total BILO	Feminino	91,87	16,877	260	3,801	≤0,001
	Masculino	84,10	16,022			

*Feminino (n = 114); Masculino (n = 148)

Observando as pesquisas anteriores com a BILOv3, foi verificado que no estudo de Freitas (2011) houve a diferenciação no desempenho em função do sexo nas provas SL ($t[99] = 2,79; p \leq 0,01$), OLV ($t[99] = 2,93; p \leq 0,01$), CF ($t[99] = 1,97; p = 0,05$), CH ($t[99] = 2,10; p = 0,04$) e no total ($t[99] = 2,71; p \leq 0,01$) com maiores pontuações para as meninas do que para os meninos do Rio Grande do Norte. Também no estudo de Joly e cols. (2010) foram encontradas diferenças estatisticamente significativas nas provas CM ($t[83,57] = 4,17; p < 0,001$) e SL ($t[119,85] = 2,37; p < 0,001$), com melhores desempenhos para as meninas paulistas de escolas privadas. No estudo de Cruz (2011), realizado no interior do estado de São Paulo, houve a diferenciação por sexo nas provas CM ($t[313] = 2,094; p = 0,037$) e no total ($t[318] = 1,9709; p = 0,05$) também com vantagem para as meninas. Contudo, nos estudos de Istome (2010) realizado em São Paulo capital e interior e no de Soares (2011), com amostra de Macapá, estas diferenças não foram encontradas. Também nas versões anteriores não foram encontradas diferenças por sexo, conforme Almeida e Joly (2008); Issa (2008) e Joly e Piovezan (2012) com a BILOv1, e Joly, Reppold e Dias (2009) e Joly, Martins, Lopes e Lemos (2009), com a BILOv2.

A leitura de todos os resultados apresentados nos estudos com BILO referentes às diferenças de desempenho em função do sexo nas diferentes regiões brasileiras, mostra que há um predomínio de melhores desempenhos para as meninas, ainda que a diferença estatisticamente significativa ocorra somente em provas específicas de acordo com a amostra investigada. Atenção deve ser dada a estas diferenças para o estabelecimento de normas separadas para cada sexo.

Como forma de verificar as diferenças nas médias das provas e do total da BILOv3 levando-se em consideração todas as características dos alunos, realizou-se uma Análise de Variância Multivariada (MANOVA). Esta análise informa se existe qualquer efeito das

variáveis independentes na combinação linear das variáveis dependentes. Para tanto, definiu-se como variáveis dependentes os escores da BILOv3 (por prova e total) e como variáveis independentes a idade, o sexo e o ano escolar dos alunos. Desta análise, o valor do F de λ de *Wilks* indicou somente o efeito estatisticamente significativo do sexo ($F[244] = 2,460; p = 0,025$) na combinação linear das variáveis dependentes. Tal efeito já foi mencionado nas análises expostas anteriormente. As MANOVAs com o Raven e com o TVIP covariando nessas mesmas variáveis também apresentaram os mesmos resultados.

Análise do tempo utilizado na execução da BILOv3

Com o objetivo de analisar o tempo utilizado pelas crianças na execução das provas da BILOv3, a Tabela 15, apresenta os tempos, em segundos de cada prova e total. É importante destacar que o tempo na BILO começa a ser registrado a partir do momento em que se finaliza o áudio contendo a consigna de cada item e termina quando a criança confirma a resposta. Por isso, o tempo apresentado nas análises não é o tempo total em que a criança manteve-se em atividade, pois este não contempla o tempo usado para o áudio. O tempo de cada prova da BILO é verificado a partir da soma do tempo gasto em cada item, bem como o tempo total da BILO é a somatória do tempo gasto nos itens de todas as provas.

Ressalta-se que na BILO o tempo é registrado em segundos o tempo no TVIP é originariamente registrado em milésimos de segundos, e nas Matrizes Progressivas Coloridas de Raven o tempo foi registrado em minutos. Na sequência estão apresentados os tempos das provas e do total da BILO, bem como, novamente são trazidos os tempos do TVIP e do Raven, contudo, todos os tempos foram convertidos em segundos para que as

unidades de medida fossem equivalentes, de modo a facilitar a visualização e comparação.

Tabela 15

Estatísticas descritivas do tempo na realização da BILO, TVIP e Raven.

Prova	Mínimo	Máximo	Média	Desvio padrão
TVIP	167,40	1417	555,22	149,590
Raven	240	1140	397,80	115,14
CM	1	199	31,19	27,852
SL	52	1051	205,31	102,356
OLV	61	974	177,01	87,905
IH	3	113	24,23	18,982
CF	1	86	19,92	13,469
CH	0	124	13,25	13,485
Total BILO	176	1662	470,92	205,393

Observa-se na Tabela 15 que, de todas as provas da BILO, a SL e a OLV foram as provas que demandaram maior tempo de execução. Isto provavelmente se dá devido ao fato de que a tarefa do respondente é organizar as cenas de cada série de modo que haja relação de causa e efeito e contiguidade entre elas, o que requer a capacidade de memória para sequência, orientação direita-esquerda e orientação temporal, além da maior habilidade de manuseio do mouse. Nas demais provas a tarefa exigia que se escolhesse uma opção dentre três, portanto, o aluno deveria clicar uma vez sobre a resposta que julgasse correta, o que justifica os menores tempos na execução da tarefa. O aparecimento de tempos ínfimos como zero, um e três segundos para a execução de uma prova inteira (CH, CM, CF e IH) pode ser devido ao fato do aluno clicar em sua escolha imediatamente depois de terminado o áudio (pois as figuras já estão expostas enquanto ele ouve a consigna), o programa não é sensível para registrar o tempo, uma vez que sua unidade de medida é o segundo,

diferentemente do TVIP que se utiliza de milissegundos. Há que se considerar também a possibilidade de erro de medida devido à velocidade de processamento do computador, que, se estiver comprometida, quando as crianças efetuam os sucessivos cliques, pode haver problemas no registro do tempo, conforme observado nos vários travamentos observados durante as aplicações.

Como forma de verificar a associação dos tempos gastos em cada instrumento, foi realizada uma análise correlacional utilizando-se o r de Pearson. Da análise observou-se moderada associação positiva ($r = 0,403$; $p < 0,001$) entre o tempo da BILOv3 ($M = 470,92$; $DP = 205,393$) e o do TVIP ($M = 555,22$; $DP = 149,590$) indicando uma tendência a alunos com maior tempo na BILOv3 também demandarem mais tempo no TVIP e os com menor tempo, demandarem menos tempo. Este dado decorre da similaridade entre os construtos, ambos relativos à linguagem, e ao processamento cognitivo que eles demandam na execução das tarefas solicitadas. Em outras palavras, os processos cognitivos que englobam a entrada da informação (percepção, atenção, memória e o componente executivo) e os processos de saída, no caso, resposta dada com o uso do *mouse* (Narbona & Fernandes, 2005; Andre & Phye, citado por Gerber, 1996). Não houve correlação estatisticamente significativa entre o tempo da BILO e do Raven ($r = -0,015$; $p = 0,813$) e entre o tempo do TVIP e do Raven houve uma fraca associação ($r = 0,197$; $p = 0,001$).

Visando verificar se existiam diferenças no tempo de execução das provas em função do ano escolar cursado, as médias foram comparadas por meio da ANOVA. Observa-se na Tabela 16 que o tempo de execução e a variabilidade de tempos diminuem à medida que os anos escolares aumentam. Exceção feita à variabilidade da prova OLV, em que um aluno com tempo elevado (974 segundos) aumentou o desvio padrão desta prova.

Tabela 16

Estatísticas descritivas dos tempos em segundos por prova e total da BILOv3 em função do ano escolar.

Prova	Ano escolar	N	Mínimo	Máximo	Média	Desvio Padrão	F[2, 259]	p
CM	Infantil	85	6	199	50,32	38,423	40,722	<0,001
	1º Ano	96	1	72	25,37	16,301		
	2º Ano	81	3	41	18,00	7,728		
	Total	262	1	199	31,19	27,852		
SL	Infantil	85	63	1051	262,45	141,941	22,982	<0,001
	1º Ano	96	52	409	175,27	60,335		
	2º Ano	81	78	393	180,96	59,592		
	Total	262	52	1051	205,31	102,356		
OLV	Infantil	85	61	507	216,72	87,738	14,270	<0,001
	1º Ano	96	66	297	160,89	52,023		
	2º Ano	81	62	974	154,46	106,823		
	Total	262	61	974	177,01	87,905		
IH	Infantil	85	7	113	40,15	23,613	67,000	<0,001
	1º Ano	96	3	56	17,46	9,808		
	2º Ano	81	4	42	15,54	8,662		
	Total	262	3	113	24,23	18,982		
CF	Infantil	85	4	86	27,91	17,071	26,624	<0,001
	1º Ano	96	2	57	16,61	9,615		
	2º Ano	81	1	42	15,47	8,669		
	Total	262	1	86	19,92	13,469		
CH	Infantil	85	3	124	19,61	18,131	15,959	<0,001
	1º Ano	96	0	71	10,95	10,915		
	2º Ano	81	0	36	9,31	6,445		
	Total	262	0	124	13,25	13,485		
Total BILO	Infantil	85	184	1662	617,15	248,932	42,029	<0,001
	1º Ano	96	176	801	406,55	117,924		
	2º Ano	81	226	1337	393,74	149,434		
	Total	262	176	1662	470,92	205,393		

As diferenças observadas nesta análise foram estatisticamente significativas, sendo que o teste *Post Hoc* utilizando-se o *Tukey HSD*, indicou que em todas as provas e no total da BILOv3 ($p \leq 0,001$), houve a formação de dois grupos, um grupo com os alunos do

Ensino Infantil e outro grupo com os alunos dos dois anos do Ensino Fundamental, diferenciando, portanto, os tempos dos alunos por nível de ensino.

Ressalta-se que a escola do Ensino Infantil não possui sala de informática e a ocasião da testagem, para a grande maioria dos alunos, foi o primeiro contato com o computador. Já os alunos do Ensino Fundamental estavam familiarizados com a informática, uma vez que, a escola possui sala de informática e aulas semanais de computação. A falta de acesso à informática pode ter sido um fator representativo do maior tempo utilizado pelas crianças do Infantil, pois esta situação aumenta a dificuldade de execução da tarefa, conforme advertido por Wall (2000).

Contudo considera-se que mesmo em crianças com acesso à informática ocorra a diferenciação do tempo em função do ano escolar, conforme observado no estudo de Istome (2010) com crianças de instituições privadas, em que houve a constatação de menores tempos para crianças do 1º ano em relação ao Infantil ($t_{\text{Total}}(155) = 3,627; p < 0,001$). Do mesmo modo, Freitas (2011) e Cruz (2012) também constataram o efeito da escolarização nos tempos.

A análise da diferença de tempo para realização dos testes em função da idade, conforme apresentado na Tabela 17, revela que ocorre o decréscimo do tempo de execução dos alunos de 5 até 8 anos em todas as provas, exceto nas idades de 7 e 8 anos nas provas OLV e CF. A ANOVA constatou que estas diferenças foram estatisticamente significativas não podendo ser atribuídas ao acaso.

Tabela 17

Estatísticas descritivas dos tempos por prova e total da BILOv3 em função da idade.

Escore	Idade (anos)	N	Mínimo	Máximo	Média	Desvio Padrão	F[3, 258]	p
CM	5	47	10	199	58,70	44,754	28,200	<0,001
	6	97	1	136	31,15	21,491		
	7	92	2	72	21,38	12,667		
	8	26	6	36	16,69	8,245		
SL	5	47	63	614	257,55	126,991	7,476	<0,001
	6	97	78	1051	212,85	118,135		
	7	92	52	409	178,90	63,742		
	8	26	107	299	177,50	54,518		
OLV	5	47	61	507	210,85	93,401	5,833	0,001
	6	97	65	433	184,35	71,626		
	7	92	62	358	150,09	53,527		
	8	26	67	974	185,04	171,634		
IH	5	47	11	103	38,40	23,314	20,330	<0,001
	6	97	3	113	27,04	20,763		
	7	92	3	56	17,34	10,066		
	8	26	4	28	12,85	5,598		
CF	5	47	6	76	28,91	17,034	13,465	<0,001
	6	97	2	86	21,11	13,864		
	7	92	1	47	15,06	8,845		
	8	26	3	42	16,65	9,239		
CH	5	47	3	74	18,89	15,093	6,890	<0,001
	6	97	0	124	15,15	16,923		
	7	92	0	46	9,55	7,672		
	8	26	2	29	9,31	5,795		
Total BILO	5	47	213	1218	613,32	252,735	15,068	<0,001
	6	97	176	1662	491,66	206,740		
	7	92	206	812	392,32	115,224		
	8	26	228	1337	418,04	214,385		

O teste *Post Hoc* usando o Tukey HSD indicou que na prova CM e IH, exceto para os alunos de 7 e 8 anos ($p = 0,808$), todas as demais idades diferiram significativamente umas das outras ($p \leq 0,05$). Na prova SL, houve somente a diferenciação entre a idade de 5 anos para com todas as outras ($p \leq 0,05$), sendo que, as idades 6, 7 e 8 anos não diferiram

estatisticamente entre si (6 e 7, $p = 0,099$; 6 e 8, $p = 0,383$; 7 e 8, $p = 1,000$). Na prova OLV a diferença ocorreu somente entre as idades de 5 e 7 ($p = 0,001$) e 6 e 7 ($p = 0,044$) anos. Na prova CF, houve a diferenciação entre os alunos com 5 anos dos alunos com 6 ($p = 0,003$), 7 ($p \leq 0,001$) e 8 ($p = 0,001$) anos. Já na prova CH a diferença se deu entre as idades de 5 e 7 ($p \leq 0,001$), 5 e 8 ($p = 0,015$) e 6 e 7 ($p = 0,019$) anos, não ocorrendo entre as idades de 5 e 6 ($p = 0,361$); entre 6 e 8 ($p = 187$), e entre 7 e 8 anos ($p = 1,000$). Por fim, a diferença observada no tempo total da BILOv3, ocorreu entre as idades de 5 anos para com todas as outras ($p \leq 0,002$), e entre 6 e 7 ($p = 0,004$) anos, não sendo consideradas estatisticamente significativas as diferenças na idade de 8 para com as de 6 ($p = 0,328$) e 7 ($p = 0,938$) anos.

Ao que pode ser observado, nas diferenças da média de tempo, assim como encontrado na média dos escores, mostram-se melhor definidas em cada ano escolar do que quando observadas em função da idade. Istome (2010) encontrou diferença entre as crianças com cinco anos e as de seis anos para a execução de todas as provas e no total da BILO ($t[155] = 3,627$; $p \leq 0,000$). No estudo de Cruz (2012) foram encontradas diferenças em relação á idade somente nas provas CM, SL, OLV, e na BILO total ($F[7, 317] = 2,558$; $p = 0,014$). Já no estudo de Freitas (2011), não houve diferença estatisticamente significativa entre o grupo etário de 5 e 6 anos para o grupo de 7 e 8 nas provas nem no total da bateria.

Visando verificar se houve diferença no tempo de execução das provas em função do sexo, por meio do teste t de *Student* compararam-se as médias dos meninos e das meninas. Observa-se na Tabela 18 que as meninas utilizaram menos tempo na execução das provas CM, IH, CF, CH e no total da BILOv3 do que os meninos. Contudo, estas diferenças se mostram estatisticamente significativas somente nas provas IH ($t[260] = 2,897$; $p = 0,004$), e CH ($t[257] = -2,027$; $p = 0,044$). No estudo de Istome (2010)

foi observada a diferença nas provas SL ($t[155] = 2,770$; $p \leq 0,006$) e OLV ($t[155] = 2,743$; $p \leq 0,007$, entretanto, com maior tempo para as meninas. Nos estudos de Freitas (2011) e de Cruz (2012) não foram encontradas diferenças estatisticamente significativas de tempo de execução em relação ao sexo. Diferentemente do encontrado com relação ao desempenho na BILOv3 em função do sexo, em relação ao tempo de execução esta diferenciação aparece como um caso isolado, características específicas desta amostra.

Tabela 18

Estatísticas descritivas dos tempos em segundos por prova e total da BILOv3 em função do sexo do aluno

Prova	Sexo	N	Média	Desvio Padrão	gl	t	p																																																																				
CM	Feminino	114	28,05	27,761	243,260	-1,604	0,110																																																																				
	Masculino	148	33,60	27,775				SL	Feminino	114	208,46	119,356	260	0,437	0,663	Masculino	148	202,89	87,390	OLV	Feminino	114	178,28	103,347	260	0,205	0,838	Masculino	148	176,03	74,212	IH	Feminino	114	20,41	16,904	260	-2,897	0,004	Masculino	148	27,17	20,002	CF	Feminino	114	18,49	14,121	260	-1,515	0,131	Masculino	148	21,03	12,884	CH	Feminino	114	11,42	10,659	257,888	-2,027	0,044	Masculino	148	14,66	15,196	Total BILO	Feminino	114	465,12	223,176	260	-0,400	0,689
SL	Feminino	114	208,46	119,356	260	0,437	0,663																																																																				
	Masculino	148	202,89	87,390				OLV	Feminino	114	178,28	103,347	260	0,205	0,838	Masculino	148	176,03	74,212	IH	Feminino	114	20,41	16,904	260	-2,897	0,004	Masculino	148	27,17	20,002	CF	Feminino	114	18,49	14,121	260	-1,515	0,131	Masculino	148	21,03	12,884	CH	Feminino	114	11,42	10,659	257,888	-2,027	0,044	Masculino	148	14,66	15,196	Total BILO	Feminino	114	465,12	223,176	260	-0,400	0,689	Masculino	148	475,38	191,230								
OLV	Feminino	114	178,28	103,347	260	0,205	0,838																																																																				
	Masculino	148	176,03	74,212				IH	Feminino	114	20,41	16,904	260	-2,897	0,004	Masculino	148	27,17	20,002	CF	Feminino	114	18,49	14,121	260	-1,515	0,131	Masculino	148	21,03	12,884	CH	Feminino	114	11,42	10,659	257,888	-2,027	0,044	Masculino	148	14,66	15,196	Total BILO	Feminino	114	465,12	223,176	260	-0,400	0,689	Masculino	148	475,38	191,230																				
IH	Feminino	114	20,41	16,904	260	-2,897	0,004																																																																				
	Masculino	148	27,17	20,002				CF	Feminino	114	18,49	14,121	260	-1,515	0,131	Masculino	148	21,03	12,884	CH	Feminino	114	11,42	10,659	257,888	-2,027	0,044	Masculino	148	14,66	15,196	Total BILO	Feminino	114	465,12	223,176	260	-0,400	0,689	Masculino	148	475,38	191,230																																
CF	Feminino	114	18,49	14,121	260	-1,515	0,131																																																																				
	Masculino	148	21,03	12,884				CH	Feminino	114	11,42	10,659	257,888	-2,027	0,044	Masculino	148	14,66	15,196	Total BILO	Feminino	114	465,12	223,176	260	-0,400	0,689	Masculino	148	475,38	191,230																																												
CH	Feminino	114	11,42	10,659	257,888	-2,027	0,044																																																																				
	Masculino	148	14,66	15,196				Total BILO	Feminino	114	465,12	223,176	260	-0,400	0,689	Masculino	148	475,38	191,230																																																								
Total BILO	Feminino	114	465,12	223,176	260	-0,400	0,689																																																																				
	Masculino	148	475,38	191,230																																																																							

4. 4 Análises Correlacionais

Como forma de verificar a existência de consistência na relação entre os escores dos três instrumentos, obtendo assim, a evidência de validade para a BILOv3 por meio da correlação com outras medidas convergentes, quais sejam, o vocabulário e a inteligência (AERA, APA & NMCE, 1999; Anastasi & Urbina, 2000) foi realizada uma análise de correlação de *Pearson*. A Tabela 19 apresenta a análise correlacional entre os escores das provas e total da BILOv3, o desempenho no Teste de Vocabulário por Imagens TVIP e o desempenho nas Matrizes Progressivas Coloridas de Raven.

Tabela 19

Correlação entre os escores da BILOv3, o Teste de Vocabulário por Imagens Peabody – TVIP e as Matrizes Progressivas Coloridas de Raven

Provas	Correlação Pearson	TVIP	Raven
CM	<i>r</i>	0,482	0,265
	<i>p</i>	< 0,001	< 0,001
SL	<i>r</i>	0,493	0,439
	<i>p</i>	< 0,001	< 0,001
OLV	<i>r</i>	0,407	0,401
	<i>p</i>	<0,001	< 0,001
IH	<i>r</i>	0,534	0,411
	<i>p</i>	< 0,001	< 0,001
CF	<i>r</i>	0,478	0,328
	<i>p</i>	< 0,001	< 0,001
CH	<i>r</i>	0,436	0,253
	<i>p</i>	< 0,001	< 0,001
Total	<i>r</i>	0,619	0,493
	<i>p</i>	< 0,001	< 0,001

De acordo com a Tabela 19, constata-se correlações positivas e altamente significativas ($p > 0,001$) entre os escores totais das provas da BILOv3 e os escores de vocabulário ($r_{\text{Total}} = 0,619$; $p > 0,001$) e de inteligência ($r_{\text{Total}} = 0,493$; $p > 0,001$), comprovando que o desempenho em compreensão da linguagem oral está associado ao desempenho em vocabulário e com a inteligência, de modo que, a medida que um aumenta os outros também aumentam, assim como quando diminuem os outros também diminuem. De acordo com Dancey e Reidy (2006), as correlações são classificadas em função de sua magnitude com a denominação Zero (correlações de 0 a 0,09), Fraca (correlações de 0,1 a 0,3), Moderada (correlações de 0,4 a 0,6); Forte (correlações de 0,7 a 0,9) e Perfeita (iguais a 1).

Com relação ao vocabulário, nota-se que as associações foram positivas e moderadas entre todas as provas ($r_{\text{CM}} = 0,482$; $r_{\text{SL}} = 0,493$; $r_{\text{OLV}} = 0,407$; $r_{\text{IH}} = 0,534$; $r_{\text{CF}} = 0,478$; $r_{\text{CH}} = 0,436$) e com o total da BILO ($r_{\text{Total}} = 0,619$). Estudos anteriores já haviam comprovado estas associações, conforme exposto na Tabela 20.

Tabela 20

Coefficientes de correlação de Pearson entre a BILO e o TVIP em estudos anteriores

	Almeida e Joly (2008)	Joly e cols.* (2009)	Istome (2010)	Cruz (2012)	Atual
	TVIP x BILOv1	TVIP x BILOv2	TVIP x BILOv3	TVIP x BILOv3	TVIP x BILOv3
CM	0,353	0,31	0,294	0,356	0,482
SL	0,394	0,33	0,333	0,490	0,493
OLV	0,400	0,50	0,394	0,487	0,407
IH	0,304	0,37	0,381	0,331	0,534
CF	0,308	0,31	0,272	0,411	0,478
CH	0,320	0,22	0,372	0,300	0,436
Total	0,431	0,42	0,478	0,563	0,619

*Dados expostos conforme os autores disponibilizaram

Da comparação com estudos anteriores, nota-se que as associações entre o vocabulário e compreensão da linguagem oral mostram-se constantes em todos os estudos, sendo que, no presente estudo as correlações obtiveram as maiores magnitudes. Estas associações são previstas uma vez que o léxico receptivo, ou vocabulário, está presente nas tarefas das provas da BILOv3, quer seja para a compreensão das consignas de cada prova, quer seja na execução do conteúdo por ela avaliado.

Em relação à prova CM, “Compreensão Morfosintática”, que envolve a avaliação da compreensão das palavras no agrupamento destas em frases (Le Normand, 2005), entende-se que a identificação lexical está totalmente imbricada nesta tarefa (Chevrie-Muller & Narbona, 2005). No que se refere à prova SL, “Sequência Lógica”, que avalia a capacidade de organizar as cenas de modo lógico-temporal, sem ter um estímulo auditivo, utilizando-se somente os estímulos visuais, entende-se que a associação desta prova com o vocabulário se deva à necessidade de compreensão das instruções auditivo-verbais para a realização da tarefa. Entende-se que se mostra imprescindível o vocabulário, uma vez que, esta foi a segunda maior associação encontrada entre o TVIP e a prova da BILOv3 ($r = 0,493$; $p < 0,001$), e a maior no estudo de Cruz (2012) ($r = 0,490$; $p < 0,001$).

No caso da prova OLV, “Organização Lógico Verbal”, que requer a mesma organização de cenas que a SL, só que com o estímulo auditivo-verbal que conta a história que as cenas representam, o envolvimento do léxico está presente tanto na compreensão das consignas quanto na compreensão da história ouvida para posterior organização. Na prova IH, que envolve responder a questões de múltipla escolha sobre uma história previamente contada, do mesmo modo, considera-se o acionamento do sistema lexical para a compreensão do texto e das perguntas que seguem (Chevrie-Muller & Narbona, 2005), nesta prova houve o maior coeficiente de correlação ($r = 0,534$; $p < 0,001$). Na Completar Frases e Completar Histórias em que o respondente é solicitado a completar a lacuna com a

figura que dê sentido à palavra, observa-se o uso do léxico, em se tratando do significado da palavra, bem como da semântica (regras gramaticais de composição de frases, orações e sentenças) e a pragmática (contextualização), conforme especificado no MNPL (Chevrie-Muller & Narbona, 2005).

No que tange à correlação entre os escores da BILOv3 e a inteligência, nota-se que as associações observadas nas provas SL ($r = 0,439$; $p < 0,001$), OLV ($r = 0,401$; $p < 0,001$) e IH ($r = 0,411$; $p < 0,001$) foram moderadas, influenciando a magnitude da associação do total ($r = 0,493$; $p < 0,001$). Já nas demais provas as associações foram de baixa magnitude ($r_{CM} = 0,265$; $r_{CF} = 0,328$; $r_{CH} = 0,253$). A associação entre os escores do TVIP e do Raven também foi positiva e moderada ($r = 0,442$; $p < 0,001$), assim como encontrado em relação à inteligência e total da compreensão da linguagem oral avaliada pela BILOv3.

Há que se ponderar que a educação de relações, avaliada pelo Raven, é um componente do “fator geral” – g – (Spearman, 1927) e, portanto, considerado um fator subjacente comum a diversas capacidades acadêmicas. Mesmo referindo-se a um processo não verbal, a capacidade edutiva está envolvida em certa medida na compreensão verbal, já que capacita compreender novas situações e estabelecer relações contextuais que permitem inferir o significado das palavras, frases, orações e textos (Angelini e cols, 1990; Eysenck, 1994). É o que se observa em relação à resolução das provas da BILOv3. No caso da SL, a prova com maior coeficiente de correlação ($r = 0,439$; $p < 0,001$), entende-se que esta seja a prova que mais envolve a capacidade de estabelecer relações, porquanto é com base na observação do conjunto das cenas embaralhadas e dos detalhes que indicam a relação de tempo, causa e efeito que o avaliando vai organizar a história.

As tarefas requeridas pela prova OLV e IH, que também obtiveram associações moderadas com o raciocínio ($r = 0,401$ e $r = 0,411$; $p < 0,001$, respectivamente), embora

envolvam mais habilidades verbais, também concentram acentuado componente de raciocínio, atenção e memória. Com relação à prova CM, de conteúdo predominantemente verbal, a capacidade edutiva estaria ligada às possíveis inferências que os alunos utilizariam para compreender vocábulos desconhecidos (assim como ocorreu no TVIP, que teve associação moderada - $r = 0,442$ – com o Raven), contudo, tendo esta prova apresentado demasiado grau de facilidade, pode não ter havido expressivo uso de inferências. O efeito teto também pode ter influenciado as fracas associações com as provas CF e CH. Isto porque, estas provas deveriam requerer operações mentais mais refinadas, conquanto englobam as capacidades pragmáticas e metalinguísticas, que estariam associadas à educação de relações.

No estudo de Joly e Piovezan (2012), quando correlacionaram a BILOv1 com a Escala de Maturidade Mental Colúmbia, que avalia o raciocínio geral, as associações estatisticamente significativas se deram somente nas provas SL ($r = 0,282$; $p = 0,005$), OLV ($r = 0,299$; $p = 0,003$), IH ($r = 0,300$; $p = 0,003$) e CH ($r = 0,215$; $p = 0,026$) e no escore total ($r = 0,292$; $p = 0,004$). Ainda que se tratem de diferentes instrumentos de medida das habilidades cognitivas observa-se nos dois estudos a maior relação entre as provas SL, OLV e IH com a inteligência.

Na sequência foram realizadas análises de correlação parciais, as quais controlaram os efeitos da idade, ano escolar, sexo e de suas combinações lineares sobre as correlações entre os escores apresentados nos três instrumentos (Dancey & Reidy, 2006). A Tabela 21 apresenta as correlações parciais entre a BILOv3 e o TVIP.

Tabela 21

*Correlação total e parcial entre os escores da BILOv3 e do Teste de Vocabulário por
Imagens Peabody – TVIP*

Provas		Correlação total TVIP	Idade	Ano escolar	Sexo	Idade *Ano	Idade *Sexo	Ano *Sexo	Ano *Idade *Sexo
CM	<i>r</i>	0,482	0,358	0,346	0,476	0,338	0,357	0,345	0,337
	<i>p</i>	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
SL	<i>r</i>	0,493	0,319	0,307	0,479	0,292	0,311	0,300	0,285
	<i>p</i>	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
OLV	<i>r</i>	0,407	0,245	0,227	0,392	0,217	0,237	0,220	0,210
	<i>p</i>	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
IH	<i>r</i>	0,534	0,435	0,410	0,521	0,413	0,429	0,405	0,409
	<i>p</i>	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
CF	<i>r</i>	0,478	0,385	0,359	0,470	0,363	0,382	0,357	0,361
	<i>p</i>	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
CH	<i>r</i>	0,436	0,376	0,344	0,430	0,356	0,374	0,342	0,354
	<i>p</i>	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
Total	<i>r</i>	0,619	0,477	0,460	0,609	0,453	0,472	0,456	0,450
	<i>p</i>	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

Ao analisar os resultados apresentados na Tabela 20, nota-se que sem o efeito da idade o valor do *r* de Pearson diminui em todas as provas e no total da BILOv3, salientado a relevância da idade na definição da magnitude da correlação. De modo ainda mais acentuado, observa-se a diminuição do grau de correlação, quando controlada a variável

“Ano escolar”, sugerindo que a mesma tem maior impacto nas correlações apresentadas. Da associação destas duas variáveis, nota-se, na coluna “Idade*Ano”, que o valor das associações diminui mais ainda, deixando claro o impacto dos critérios desenvolvimentais na relação da compreensão da linguagem oral com o vocabulário receptivo.

A variável “sexo” é a que menos interfere na correlação, uma vez que, com a sua influência as correlações pouco diminuem, praticamente mantendo-se com moderada intensidade. Quando da junção da variável “Sexo” com as demais, conforme colunas “Idade*Sexo”, “Ano*Sexo” e “Idade*Ano*Sexo” observa-se aumento da influência sobre a diminuição da correlação, decorrentes da influência já descrita da escolaridade e do ano escolar sobre as correlações. Em todas as provas as correlações tornaram-se fracas, contudo o total da BILOv3 manteve-se moderadamente correlacionado ao desempenho no vocabulário. Igualmente, no estudo de Cruz (2012), as correlações parciais também demonstraram a influência da escolaridade e da idade na associação da BILOv3 com o TVIP.

Também foi realizada a análise de correlação parcial, controlando os efeitos da idade, ano escolar, sexo e de suas combinações lineares sobre as correlações dos escores da BILOv3 com os das Matrizes Progressivas Coloridas de Raven. A Tabela 22 apresenta os coeficientes de correlação desta análise.

Tabela 22

Correlação total e parcial entre os escores da BILOv3 e das Matrizes Progressivas

Coloridas de Raven

Provas		Correlação total Raven	Idade	Ano Escolar	Sexo	Idade *Ano	Idade *Sexo	Ano *Sexo	Ano *Idade *Sexo
CM	<i>r</i>	0,265	0,119	0,103	0,262	0,093	0,120	0,104	0,094
	<i>p</i>	< 0,001	0,054	0,049	< 0,001	0,066	0,027	0,048	0,065
SL	<i>r</i>	0,439	0,288	0,273	0,442	0,261	0,297	0,283	0,271
	<i>p</i>	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
OLV	<i>r</i>	0,401	0,269	0,251	0,401	0,244	0,275	0,258	0,251
	<i>p</i>	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
IH	<i>r</i>	0,411	0,304	0,280	0,412	0,279	0,312	0,289	0,289
	<i>p</i>	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
CF	<i>r</i>	0,328	0,222	0,197	0,326	0,197	0,224	0,199	0,199
	<i>p</i>	< 0,001	< 0,001	0,001	< 0,001	0,001	< 0,001	0,001	0,001
CH	<i>r</i>	0,253	0,172	0,142	0,251	0,147	0,173	0,143	0,148
	<i>p</i>	< 0,001	0,005	0,011	< 0,001	0,009	0,003	0,011	0,009
Total	<i>r</i>	0,493	0,347	0,326	0,496	0,318	0,356	0,336	0,328
	<i>p</i>	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001

Novamente constata-se o maior impacto da variável “Ano escolar” sobre o *r* de Pearson em todas as provas e total da BILOv3, seguida pela variável idade. Da associação destas duas variáveis, na coluna Idade*Ano, observa-se o menor grau de correlação, diminuindo ao ponto de tornar a correlação da prova CM nula (Dancey e Reidy, 2006) e sem significância estatística ($\geq 0,05$), ou seja, podendo ser atribuída ao acaso.

Já em relação à variável “Sexo”, sem a sua interferência, a correlação entre compreensão da linguagem oral e inteligência torna-se ligeiramente mais forte no total da

BILOv3 e nas provas SL e IH, mantendo o mesmo valor de r na prova OLV e tendo uma leve diminuição nas provas CM, CP e CH. Estes dados indicam que esta variável não se mostrou determinante do grau de correlação encontrada nesta análise.

Análise das diferenças por grupos extremos

Com o intuito de verificar se a BILOv3 discriminava o desempenho dos participantes em função de sua amplitude de vocabulário avaliada pelo TVIP, foi realizada uma análise das diferenças por grupos extremos. As estatísticas descritivas e as análises de diferença de média constam na Tabela 23.

Tabela 23

Estatísticas descritivas do desempenho na BILOv3 em função dos grupos extremos

formados pelo TVIP

Provas	Grupos Extremos TVIP	N	Média	Desvio Padrão	gl	t	p
CM	Inferior	68	15,65	3,676	77,871	-7,750	>0,001
	Superior	75	19,24	1,101			
SL	Inferior	68	3,59	2,201	93,151	-10,982	>0,001
	Superior	75	12,15	6,341			
OLV	Inferior	68	5,40	2,227	94,396	-8,655	>0,001
	Superior	75	12,03	6,208			
IH	Inferior	68	14,00	3,628	97,026	-10,127	>0,001
	Superior	75	18,95	1,837			
CF	Inferior	68	15,43	3,320	86,864	-7,750	>0,001
	Superior	75	18,77	1,351			
CH	Inferior	68	17,21	2,899	85,308	-6,063	>0,001
	Superior	75	19,48	1,131			
Total BILO	Inferior	68	71,26	11,574	140,204	-13,832	>0,001
	Superior	75	100,61	13,781			

Os participantes foram divididos em dois grupos o “GPI” que contava com aqueles (N=68) cujo escore foi inferior a 55 pontos no TVIP (percentil ≤ 25) e, portanto tinham menor escore no vocabulário receptivo, e o “GPS” constituído pelos alunos (N=76) com escore no TVIP superior a 69 pontos (percentil ≥ 75), que revelavam vocabulário mais amplo. É possível observar na Tabela 22 que o desempenho do grupo com menor vocabulário (GPI) é também inferior na compreensão em linguagem oral quando observadas as médias por prova e no total da BILOv3 ($M_{GPI\ Total} = 71,26$; $DP = 11,57$; $M_{GPS\ Total} = 100,61$; $DP = 13,78$).

Há que se ressaltar a diferença observada na prova SL, em que a média do GPS ($M = 12,15$; $DP = 6,34$); é três vezes maior do que a do GPI ($M = 3,59$; $DP = 2,20$); e na prova OLV, em que a média do GPS ($M = 12,03$; $DP = 6,21$) equivale ao dobro da média do GPI ($M = 5,40$; $DP = 2,22$). Por outro lado, para as provas CM, IH, CP e CH os o grupo de extrema pontuação rebaixada teve pontuações próximas às dos alunos com extrema pontuação elevada, novamente denotando o efeito teto desta provas. O teste *t* de Student aplicado na diferença de médias entre o GPI e o GPS revelou que as diferenças foram estatisticamente significativas entre os grupos em todas as provas e no total da BILOv3 ($t[140] = -13,832$, $p > 0,001$), conforme encontrado também em Joly e Almeida (2008).

Do mesmo modo como realizado com o TVIP, buscou-se verificar se a BILOv3 discriminava o desempenho dos alunos em função do escore de inteligência avaliado pelas Matrizes Progressivas Coloridas de Raven, conforme Tabela 24. Foram, portanto, formados o GPI, composto pelos alunos com escore inferior a 15 pontos (N = 88) e o GPS, com os alunos com escore superior a 25 pontos (N = 76), equivalendo aos desempenhos abaixo do percentil 25 e acima do percentil 75, respectivamente.

Tabela 24

Estatísticas descritivas do desempenho na BILOv3 em função dos grupos extremos formados com base nos escores de inteligência.

Provas	Grupos Extremos Raven	N	Média	Desvio Padrão	gl	<i>t</i>	<i>p</i>
CM	Inferior	88	17,26	3,087	159,441	-3,288	≥0,001
	Superior	76	18,66	2,341			
SL	Inferior	88	5,59	4,650	126,768	-5,922	>0,001
	Superior	76	11,20	7,031			
OLV	Inferior	88	6,58	3,704	115,344	-5,985	>0,001
	Superior	76	11,62	6,483			
IH	Inferior	88	15,44	3,988	140,813	-5,934	>0,001
	Superior	76	18,39	2,251			
CF	Inferior	88	16,63	2,988	141,096	-5,062	>0,001
	Superior	76	18,51	1,693			
CH	Inferior	88	18,17	2,583	138,318	-3,463	≥0,001
	Superior	76	19,28	1,410			
Total BILO	Inferior	88	79,67	14,999	152,000	-7,196	>0,001
	Superior	76	97,66	16,750			

Dos resultados expressos na Tabela 24 constata-se que a média do GPI foi menor do que a do GPS nas provas e no total da BILOv3 ($M_{GPI\ Total} = 79,67$; $DP = 14,99$; $M_{GPS\ Total} = 97,66$; $DP = 16,75$). Novamente se observam grandes diferenças entre os grupos nas provas SL e OLV, bem como pequenas diferenças nas provas CM, IH, CF e CH. O teste *t* de Student aplicado na diferença de médias entre o GPI e o GPS revelou que as diferenças foram estatisticamente significativas entre os grupos em todas as provas e no total da BILOv3 ($t[152] = -7,196$, $p > 0,001$), indicando que a BILOv3 foi sensível para discriminar alunos com altos escores de inteligência dos alunos com escores abaixo da média. Isto posto confirmam-se as evidências de validade de critério por grupo extremo para a BILO, conforme especificam a AERA, APA e NCME (1999).

5. CONSIDERAÇÕES FINAIS

Para além de alcançar o objetivo deste estudo, qual seja, buscar evidências de validade para a BILOv3 por meio da correlação com vocabulário e inteligência, pretendeu-se traçar um paralelo entre a atual pesquisa e as pesquisas realizadas desde 2006 com a BILO, de modo a dar uma visão geral do instrumento. É notório que, desde a primeira versão, já vem sendo encontradas evidências de validade para a BILO, que atestam que a bateria apresenta qualidades psicométricas, assim como foi observado no presente estudo.

Há que se considerar a representatividade da amostra do presente estudo, que contemplou mais de 50% da população total de estudantes matriculados no último ano da Educação Infantil e 1º e 2º anos do Ensino Fundamental da cidade onde se realizou a pesquisa. Além disso, estes alunos, conforme demonstrado pelos escores do Raven e do TVIP, evidenciaram boas condições cognitivas e vocabulário adequado para a idade, tornando os resultados do presente estudo muito representativos.

Com relação ao objetivo principal, de correlacionar os escores da BILOv3 com o vocabulário e com a inteligência, as associações moderadas da BILOv3 com o TVIP atestaram, mais uma vez, a relação da compreensão da linguagem oral com o vocabulário. No que tange a associação da BILO com as Matrizes Progressivas de Raven, a associação moderada com as provas SL, OLV e IH também garantiram evidência de validade, porquanto, demonstraram empiricamente o envolvimento do raciocínio com as tarefas exigidas nestas provas. Além disso, a BILOv3 foi sensível para diferenciar os grupos extremos, formados pelos alunos com menores e alunos com maiores pontuações em

vocabulário e inteligência, o que lhe reafirmam as evidências de validade (AERA, APA & NMCE, 1999; Anastasi & Urbina, 2000).

Outro objetivo, o de buscar evidências de validade teste-critério, tendo a idade e a escolaridade como medidas, também foi alcançado, conquanto, a BILOv3 novamente mostrou-se capaz de diferenciar alunos mais novos dos mais velhos e em anos escolares mais avançados dos anos iniciais. Isto porque, o desenvolvimento psicolinguístico está associado à maturação neurológica que ocorre com o avanço da idade (fator biológico), juntamente com os estímulos ambientais, isto é, influências de ordem social e educacional (Narbona, & Fernández, 2005; Pedroso e Rotta, 2006). Sobre estes aspectos, tanto quanto foi observado no presente estudo, os estudos anteriores, também evidenciaram maior impacto da escolarização do que da idade na diferenciação de médias em compreensão da linguagem oral. Entende-se que a escolarização, fonte principal do conhecimento formal, tem, ou deveria ter, grande impacto sobre a compreensão da linguagem oral e do vocabulário da criança.

Ainda que o tempo de execução não esteja sendo aferido de modo preciso – pois há o registro de tempos “0”, a análise dele decorrente, também foi um indicador de evidência de validade, uma vez que diferenciou o tempo dos alunos mais novos do tempo dos alunos mais velhos. Sendo que, na medida em que a idade e os anos escolares aumentam, diminuem os tempos de execução das provas. Estes resultados atribuem à BILO validade por relação com critério desenvolvimental (AERA, APA & NMCE, 1999; Anastasi & Urbina, 2000). No entanto, sugere-se que, diante da importância da medida do tempo enquanto habilidade cognitiva, seja alterado o programa afim de que registre o tempo em milésimos de segundos, garantindo assim, maior aferição da latência entre o estímulo e a resposta, como ocorre no TVIP.

Em adição, observou-se a diferença significativa nas médias das provas SL, IH e no TVIP em relação ao sexo do aluno, indicando que as meninas obtiveram médias maiores do que a dos meninos. Nos estudos anteriores também foram observadas diferenças entre meninas e meninos, ainda que a diferença estatisticamente significativa ocorra somente em provas específicas de acordo com a amostra investigada. Apesar de não haver dados que diferenciem meninos de meninas no que concerne ao desenvolvimento da linguagem oral, há que se considerar que nas crianças pré-escolares há maior incidência de transtornos na linguagem sobre os meninos, na proporção de 2 a 4 para cada menina (Pedroso e Rotta, 2006). Ressalta-se a importância de tais dados quando da normatização do instrumento, uma vez que, se forem comprovadas diferenças entre os sexos, normas distintas deverão ser criadas (Urbina, 2007)

Perpassando todos esses resultados, o que se observou foi a disparidade das provas SL e OLV em relação às demais provas da bateria. O efeito teto observado nas provas CM, IH, CF e CH em contraste com as baixas pontuações da SL e OLV podem prejudicar a aferição da linguagem oral por meio da BILOv3. Tendo em vista que avaliação e o diagnóstico em linguagem oral devam ser baseados em critérios que auxiliem os profissionais a caracterizarem o desempenho e identificar os aspectos deficitários do avaliando (Chevrie-Muller & Narbona, 2005), a identificação das áreas específicas avaliadas pelas provas CM, CF e CH podem não estar sendo efetivas para as idades e anos escolares abarcados pelo Ensino Fundamental. Acrescente-se a esta análise o fato de que os menores índices de precisão da BILOv3 ocorrem nestas provas, conforme apresentados no estudo de Freitas (2011) que indicou CM = 0,68; CF = 0,65 e CH = 0,43 e no estudo de Cruz (2012) com CM = 0,71; CF = 0,55 e CH = 0,54.

Já as provas SL e OLV, apesar de apresentarem as menores médias, mostraram-se mais efetivas na diferenciação dos grupos extremos formados a partir dos piores e dos

melhores desempenhos no vocabulário em inteligência, também foram as que melhor diferenciaram as pontuações dos alunos em relação aos critérios desenvolvimentais utilizados, mostrando-se adequadas aos alunos do Ensino Fundamental. Freitas (2011) encontrou índices de precisão iguais a 0,87 para a SL e de 0,89 para a OLV. Cruz (2012) identificou alfas de Cronbah para a SL igual a 0,85 e para a OLV, 0,86.

Sugere-se que, as futuras pesquisas com a BILO se concentrem no Ensino Infantil englobando crianças desde o ingresso no ensino infantil, com amostras diferenciadas por regiões brasileiras. Contudo, adverte-se que, com relação à provas SL, deve ser investigado se as crianças estão compreendendo as instruções dadas, uma vez que, a correlação moderada encontrada entre esta prova e o vocabulário, comprovou que o nível lexical das crianças vai determinar se elas vão entender ou não as consignas da prova. Na prova SL os alunos são orientados pelo *software* a organizar a seqüência de figuras formando histórias de modo que cada história apresente começo, meio e fim. Já na prova de Organização Lógico-Verbal os alunos são instruídos a ouvir as sentenças e depois organizar as cenas com começo, meio e fim de acordo com o que ouviram. Por isso, as provas SL e OLV requerem a capacidade de memória para seqüência, de orientação direita-esquerda e orientação temporal, além da maior habilidade de manuseio do mouse, o que pode ser um dificultador para as crianças em fase de desenvolvimento abaixo dos 5 anos.

Diante das constatações do presente estudo torna-se imprescindível a análise fatorial com informação plena, bem como uma análise pela Teoria de Resposta ao Item para avaliar a terceira versão da BILO, de modo a averiguar mais profundamente os itens da bateria e, assim, orientar as alterações necessárias, conforme já indicado em Istome (2010). A partir destas análises, além dos itens, indica-se a correção do software, no que tange ao registro do tempo – de segundos para milissegundos - e na garantia de que a prova CH não trave com os repetidos acionamentos do mouse sobre o item, como foi observado nas aplicações.

Nota-se que a bateria, ao longo das três versões manteve-se constante em diferenciar as médias por critérios desenvolvimentais e também delimitou empiricamente o construto que avalia, quando da realização das análises correlacionais convergentes e divergentes as quais foi submetida. Contudo, constatou-se por meio das análises realizadas no presente estudo que a BILO_{v3}, apesar de apresentar evidências de validade que lhe atribuem características psicométricas, pode estar desperdiçando oportunidades de avaliar mais precisamente as habilidades requeridas nas provas CM, CF e CH.

Por fim, não se pode deixar de mencionar novamente que a BILO tem se mostrado, desde os estudos com as primeiras versões, como um instrumento diferenciado de avaliação, porquanto os recursos multimídia empregados na sua estrutura permitem, por um lado mensuração das habilidades cognitivas, e por outro, maior motivação e interesse das crianças em executar as tarefas, tornando a testagem uma atividade gratificante tanto para o avaliador quanto para o avaliado, conforme já referido em Istome (2010). Por isso, considera-se que, com as alterações solicitadas, o instrumento poderá ser utilizado de modo eficiente na avaliação e detecção precoce de distúrbios na linguagem receptiva.

5. REFERÊNCIAS

Acosta, V. M., Moreno, A., Ramos, V., Quintana, A., & Spino, O. (2003). *Avaliação da linguagem: teoria e prática do processamento de avaliação do comportamento lingüístico infantil*. São Paulo: Livraria Santos.

Adánez, G. A. (1999). Procedimientos de construcción y análisis de tests psicometricos. Em S. M. Wechesler & R. S. L. Guzzo (Orgs.) *Avaliação Psicológica: Perspectiva Internacional* (pp.57-100). São Paulo: Casa do Psicólogo.

AERA, APA & NMCE – American Educational Research Association, American Psychological Association & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. New York: American Educational Research Association.

Almeida, A. R.(2008). *Estudo de validade da Bateria Informatizada de Linguagem Oral* (BILO). Dissertação de Mestrado, Programa de Pós-Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba.

Almeida, A. R., & Joly, M. C. R. A. (2008). Estudo correlacional entre a Bateria Informatizada de Linguagem Oral (BILO) e Peabody. Em: XIII Conferência Internacional em Avaliação Psicológica: Formas e Contextos, Braga. Actas do XIII Conferência Internacional em Avaliação Psicológica: Formas e Contextos. Braga : Psiquilíbrios.

Alves, I. C. B., & Duarte J. L. M. (1994). *Escala de Maturidade Mental de Colúmbia – Padronização Brasileira*. São Paulo: Casa do Psicólogo.

American Speech-Language-Hearing Associationm – ASHA (2007, abril). *What is Language? What is speech?* [On-line]. Recuperado: 13/04/2010. Disponível: http://www.asha.org/public/speech/development/language_speech.htm

Anastasi, A., & Urbina, S. (2000). *Testagem Psicológica*. 7ª ed. Porto Alegre: Artes Médicas.

Angelini, A. L., Alves, I. C. B., Custódio, E. M., Duarte, W. F. & Duarte, J. L. M. (1992). *Manual Matrizes Progressivas Coloridas de Raven. Escala Especial*. São Paulo: Centro Editor de Testes e Pesquisas em Psicologia.

Bandeira, D. R., Alves, I. C. B., Giacomel, A. E., & Lorenzatto, L. (2004). Matrizes progressivas coloridas de Raven – escala especial: normas para Porto Alegre, RS. *Psicologia em Estudo*, Maringá, 9(3), 479-486.

Capovilla, A. G. S. e Dias, N. M. (2008). Habilidades de linguagem oral e sua contribuição para a posterior aquisição de leitura. *Psic*, 9(2), 135-144.

Capovilla, A. S. G., Joly, M. C. R. A., & Tonelloto, J. M. de F. (2006). Avaliação neuropsicológica e aprendizagem. Em A. P. P. Noronha, A. A. A. dos Santos, & F. F. Sisto. *Facetas do fazer em Avaliação Psicológica* (pp. 141 – 162). São Paulo: Vetor Editora.

Capovilla, F. C., & Capovilla, A. G. S. (1997). Desenvolvimento linguístico na criança dos dois aos seis anos: Tradução e standardização do Peabody Picture Vocabulary Test de Dunn & Dunn e da Language Development Survey de Rescorla. *Ciência Cognitiva*, 1(1), 353-380.

Capovilla, F. C., Nunes, L., Araújo, I., Nogueira, D., & Bernat, A. B. (1997). Versão brasileira do Teste de Vocabulário Por Imagens Peabody. *Distúrbios da Comunicação*, 8(2), 151-162.

Capovilla, F. C., & Prudêncio, E. R. (2006). Teste de vocabulário auditivo por figuras: normatização e validação preliminares. *Avaliação Psicológica*, 5(2), 189-20.

Chevrie-Müller, C., & Narbona, J. (2005). Semiologia dos distúrbios da linguagem na criança. Em: C. Chevrie-Müller & J. Narbona (Eds.), *A linguagem da criança: aspectos normais e patológicos* (pp. 52-69). Porto Alegre, RS: Artmed.

Cruz, T. P. B. P. (2011). *Estudo Psicométrico da Bateria Informatizada de Linguagem Oral (BILOv3) com Teste de Vocabulário Peabody para crianças*. Dissertação de Mestrado, Programa de Pós-Graduação *Stricto-Sensu* em Psicologia, Universidade São Francisco, Itatiba.

Dancey, C. P., & Reidy, J. (2006). *Estatística sem matemática para psicologia*. Porto Alegre: Artmed.

DocKrell, J. & McShane, J. (1993). Dificuldades específicas com linguagem. Em J. Dockrell & J. McShane, *Crianças com dificuldades de aprendizagem - uma abordagem cognitiva* (pp 57 – 83). Porto Alegre, RS: Artmed.

Dunn, L. M., Dunn, D., Capovilla, F. C., & Capovilla, A. G. S. (no prelo). Teste de Vocabulário por Figuras Peabody – Versão Brasileira. São Paulo, SP: Casa do Psicólogo.

Eysenck, M. W. (1994). *Psicologia cognitiva: um manual introdutório*. Porto Alegre: Artmed.

Freitas, J. C. P. (2011). *Estudo correlacional da Bateria Informatizada de Linguagem Oral (BILOv3) com Teste Token*. Dissertação de Mestrado, Programa de Pós Graduação *Stricto Sensu* em Psicologia, Universidade São Francisco, Itatiba.

Gerber, A. (1996). Cognição e processamento de informações. In A. Gerber e cols. *Problemas de aprendizagem relacionados à linguagem: sua natureza e tratamento* (S. Costa, trad.). Porto Alegre: Artes Médicas. (Trabalho original publicado em 1993).

Gil, R. (2002). *Neuropsicologia*. (Trad. Maria Alice Araripe Smaprio Doria.) São Paulo: Livraria Santos Editora.

Gómez, L. A., Duarte, A. M., Merchán, V., Aguirre D. C., & Pineda, D. A. (2007). Conciencia fonológica y comportamiento Verbal en niños con dificultades de Aprendizaje. *Universitas psychologica*, 6(3), 571-580.

Gurgel, L. G., Plentz, R. D. D. M., Joly, M. C. R., & Reppold, C. T. (2010). Instrumentos de avaliação da compreensão de linguagem oral em crianças e adolescentes: uma revisão sistemática da literatura. *Revista Neuropsicologia Latinoamericana*, 2(1), 1-10.

Issa, G. M. P. (2008). *Estudos de Evidências de Validade da Bateria Informatizada da Linguagem Oral – BILO*. Dissertação de Mestrado, Programa de Pós-Graduação Stricto Sensu em Psicologia, Universidade São Francisco, Itatiba, São Paulo.

Istome, A. C. (2010). *Bateria Informatizada de Linguagem Oral (versão 3): Características psicométricas para educação infantil e ensino fundamental*. Trabalho de Conclusão de Curso, do curso de Psicologia, Universidade São Francisco, Itatiba.

International Test Commission (2000). Diretrizes para o uso de testes: International Test Commission. [On-line]. Disponível: <http://www.intestcom.org/Guidelines/Test+Use.php>
Recuperado: 16 nov. 2010.

Instituto Nacional de Educação e Pesquisa Anísio Teixeira – INEP (2008) *Índice de Desenvolvimento da Educação Básica – IDB*. Disponível em <http://www.portalideb.com.br/cidade/2013-pinhalzinho/ideb>, Recuperado em, 12 de janeiro de 2013.

Joly, M. C. R. A. (2009). *Bateria Informatizada de Linguagem Oral – BILOv3*. Teste em desenvolvimento.

Joly, M. C. R. A., & Dias, A. S. (2009). Evidências de Validade de uma prova informatizada de Linguagem Oral – BILOv1. *Psicologia: Teoria e Prática*, 11(2), 50-86.

Joly, M. C. R. A., & Noronha, A. P. P. (2006). Reflexões sobre a construção de instrumentos psicológicos informatizados. Em A. P. P. Noronha, A. A. A. dos Santos & F. F. Sisto (org.). *Facetas do fazer em avaliação psicológica*. (pp.95-105). São Paulo: Vetor.

Joly, M. C. R. A., Cruz, T. P. B. P., Lima, L. B. V., Prates, E. A. R., & Leme, E. M. (2010). Avaliação informatizada da compreensão oral no ensino fundamental público e privado. *Revista de Psicologia – Universidad César Valejjo*, 12, 121-140.

Joly, M. C. R. A., Martins, D. F., Lopes, R. M. M., & Lemos, T. H. (2009). Estudo Correlacional entre a Bateria Informatizada de Linguagem Oral (BILOv2) e Peabody. Pôster apresentado no XXXII Congresso Iberoamericano de Diagnóstico e Avaliação Psicológica – Buenos Aires - Argentina – Buenos Aires – Argentina.

Joly, M. C. R. A., Martins, R. X., Abreu, M. C., Souza, P. R. R., & Cozza, H. F. P. (2004). Análise da produção científica em avaliação psicológica informatizada. *Avaliação Psicológica*, 3(2), 121-129.

Joly, M. C. R. A., Martins, R. X., Zani, A. C., Istome, A. C., & Santos, C. R. O. A. (2008). Bateria Informatizada de Linguagem Oral (BILO): validade de construto. In: XIII Conferência Internacional em Avaliação Psicológica: Formas e Contextos, Braga. Actas do XIII Conferência Internacional em Avaliação Psicológica: Formas e Contextos. Braga: Psiquilíbrios.

Joly, M. C. R. A. & Piovezan, N. M. (2012). Estudo de validade correlacional e de critério da Bateria Informatizada de Linguagem Oral (BILO) com prova de raciocínio. *Estudos de*

Psicologia (Campinas),29,499-508.

Joly, M. C. R. A., Piovezan, N. M., Soares, C. A., Lopes, R. M. M., & Martins, D. F. (2009). Avaliação das características psicométricas da bateria informatizada de linguagem oral - *BILOv2*. Pôster apresentado no III Congresso Latino Americano de Psicologia – Ulapsi, Cidade do México, México.

Joly, M. C. R. A., Reppold, C. T., & Dias, A. S. (2009). Avaliação da linguagem oral de crianças paulistas e gaúchas pela Bateria Informatizada de Linguagem Oral (BILOv2). Em C. Hutz (Org.), *Avaliação psicológica de crianças e adolescentes* (pp. 175-208). São Paulo: Casa do Psicólogo.

Joly, M. C. R. A., Welter, G. M. R., Martins, R. X., Silva, J. M., Montiel, J. M., Lopes, F., & Carvalho, M. R. (2005). Sistema de avaliação para testes informatizados (SAPI): estudo preliminar. *Psic*,6(2), 51-60.

Kaufman, D. (1996). A natureza da linguagem e sua aquisição. Em A. Gerber e cols. *Problemas de aprendizagem relacionados à linguagem: sua natureza e tratamento* (S. Costa, trad.). Porto Alegre: Artes Médicas. (Trabalho original publicado em 1993).

Le Normand, T. M. (2005). Modelos psicolinguísticos do desenvolvimento da linguagem. Em: C. Chevie-Müller & J. Narbona (Eds.), *A linguagem da criança: aspectos normais e patológicos* (pp. 52-69). Porto Alegre, RS: Artmed.

Macedo, E. C.; Capovilla, F. C.; Duduchi, M.; D’Antino, M. E. F.; & Firmo, L. S. (2006). Avaliando linguagem receptiva via teste de vocabulário por imagens *peabody*: versão tradicional *versus* computadorizada. *Psicologia: Teoria e Prática* 8(2), 40-5.

Pasquali, L.; Wechsler, S. M. & Bensusan, E. (2002). Matrizes Progressivas do Raven Infantil: um estudo de validação para o Brasil. *Avaliação Psicológica*, 1(2), 95-110.

Pedroso, F. S., & Rotta, N. T. Transtornos da linguagem. Em: N. T. Rotta; L. Ohlweiler & R. S. Riesgo (Eds), Transtornos da aprendizagem: abordagem neurobiológica e multidisciplinar (pp. 131 – 150). Porto alegre: Artmed.

Prieto, M. G., Salazar, D. P., Rebolledo, R. C., Vásquez, J. P. L., Villamizar, N. M, Pérez, A. J, & Ramírez, A. A. (2008). Trastorno Específico del Desarrollo del Lenguaje en una población infantil colombiana. *Universitas Psychologica*, 7(2), 557-569.

Reynolds, C. R., Livingston, R. B., & Wilson, V. (2009). Introduction to educational assessment. In *Measurement an assessment in Education*, ed. 2. New jersey: Person.

Spearman, C. (1927). *Las Habilidades del Hombre: su naturaleza y medición*. Buenos Aires: Paidós.

Soares, C. A. (2011). *Evidências de Validade da Bateria Informatizada de Linguagem Oral (BILOv3) em estudantes de Macapá*. Dissertação de Mestrado. Programa de Pós-Graduação *Strito-Sensu* em Psicologia, Universidade São Francisco, 2011.

Strauss, E, Sherman, E. M. S. & Spreen, O. (2006). *A compendium of neuropsychological test: administration, norms, and commentary*. (3ª ed). New York: Oxford University Press.

Trevisan, B. T.; Montiel, J. M.; Dias, N. M. & Capovilla, A. G. S. (2008). Avaliação do vocabulário receptivo em crianças do ensino fundamental: Teste de Vocabulário por Imagens Peabody (TVIP). Em A. L. Sennyey; F.C. Capovilla & J. M. Montiel (Orgs). *Transtornos de aprendizagem da avaliação à reabilitação* (361-368). São Paulo: Artes Médicas.

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

Wall, J (2000). *Technology-Delivered Assessment: Diamonds or Rocks?* [on-line] Acessado em: 22/09/2011. Disponível em: <http://www.ericdigests.org/2001-3/rocks.htm>.