

Capovilla, F.C.; **Nunes, L.R.**; Nunes, D.R.; Araújo, I.; Nogueira, D.; Bernat, A.B. Capovilla, A.G (1997). Desenvolvimento do vocabulário receptivo-auditivo da pré-escola à oitava série: Tabelas de normatização fluminense baseadas e aplicação coletiva em sala de aula da tradução brasileira do *Peabody Picture Vocabulary Test*. *Ciência Cognitiva: Teoria, Pesquisa e Aplicação,1(1)*, 371-430. (ISSN 1415-1472)

O desenvolvimento do vocabulário receptivo-auditivo da pré-escola à oitava série (6a-14a): Tábuas de normatização fluminense com base na aplicação coletiva em sala de aula da versão brasileira do *Teste de Vocabulário por Imagens Peabody*

Fernando C. Capovilla ¹, Leila R.O.P. Nunes ², Débora Nunes ³, Ivânia Araújo ³,
Daniel Nogueira ², Ana Beatriz Bernat ³, Alessandra G.S. Capovilla ⁴

¹ Instituto de Psicologia, Universidade de São Paulo, e Pesquisador Visitante da
Universidade do Estado do Rio de Janeiro

² Mestrado em Educação, Universidade do Estado do Rio de Janeiro

³ Instituto de Psicologia, Universidade Federal do Rio de Janeiro

⁴ Instituto de Psicologia, Universidade de São Paulo

Resumo

Neste estudo o *Teste de Vocabulário por Imagens Peabody (TVIP)* foi traduzido a partir da versão hispano-americana e aplicado coletivamente a uma amostra de 687 respondentes da Escola de Aplicação da Universidade do Estado do Rio de Janeiro. A amostra cobriu todos os respondentes de pré-escola até a oitava série daquela escola, com idades variando de seis a 18 anos. O desenvolvimento do vocabulário receptivo mostrou-se função direta monotônica da faixa etária e do nível de escolaridade. Foram comparados dois critérios para a avaliação das respostas: a *pontuação piso-teto* tradicional, própria para aplicação individual, e a *pontuação simples*, não-tradicional e mais apropriada para aplicação coletiva em sala de aula. A pontuação piso-teto era estabelecida de acordo com a idade e o padrão de erros da criança, avaliado conforme critérios pouco práticos. Já a pontuação simples consistia meramente no total de acertos, e foi proposta aqui por ser mais apropriada para

aplicação quando as pranchas do teste ainda não se encontram ordenadas de acordo com o grau de dificuldade para a cultura à qual se está adaptando o teste. Assim, neste estudo preliminar o mesmo desempenho foi avaliado de acordo com os dois critérios com vistas a determinar se a pontuação simples seria tão discriminativa quanto a piso-teto. A resposta foi positiva. A comparação entre os critérios foi feita por meio de análises estatísticas de comparação entre pares (e.g., um e outro nível escolar, uma e outra idade) subsequentes a ANOVAs. Assim, para a presente tradução do teste neste estudo de aplicação coletiva, a pontuação simples foi tão discriminativa quanto a piso-teto. Na composição da amostra foi observada uma correlação positiva entre o nível de escolaridade dos respondentes e o nível socio-econômico de seus pais. A contribuição do nível sócio-econômico foi controlada por meio de ANCOVA, e foi observado que o nível de escolaridade paterno (mas não materno) contribuiu para o desempenho no teste. Erro padrão de mensuração e coeficientes de consistência interna foram obtidos e revelaram que a presente tradução da adaptação hispano-americana é surpreendentemente confiável, mesmo antes de qualquer reordenação de itens a partir da análise do grau de dificuldade dos itens. Tal reordenação foi então empreendida com base na análise de itens dos dados da amostra toda, e encontra-se aqui. Finalmente, a partir das pontuações simples e piso-teto obtidas da aplicação coletiva, foram geradas tábuas de normatização, as quais encontram-se apresentadas tanto em função de nível escolar quanto de faixa etária. Este primeiro estudo objetiva permitir elaborar a adaptação brasileira definitiva do *TVIP*.

Palavras-chave: *Teste de Vocabulário por Imagens Peabody*, Vocabulário, Psicometria

The development of receptive-auditory vocabulary from pre-school to eight grade (6y-14y) in Rio de Janeiro: Normalization boards from classroom application of the Brazilian version of the Peabody Picture Vocabulary Test

Abstract

In the present study, the Hispanic-American version of the Peabody Picture Vocabulary Test (PPVT) was translated to Portuguese and applied collectively to a sample of 687 children from a public school of Rio de Janeiro city. The sample covered all the children from pre-school to eighth grade, with ages varying from six to 18 years (children from 14 to 18 were delayed). The receptive-vocabulary development was a positive monotonic function of age and school levels. Two scoring criteria were compared: ceiling-floor score (appropriate for individual assessment in the clinic) and simple score (appropriate for collective assessment in the classroom). Scoring based on ceiling-floor criteria was age and error rate according to fairly cumbersome criteria. Scoring based on simple count was based on total frequency of correct responses. It is more appropriate for assessment when the testing trials are not yet reordered according to the degree of difficulty for a given culture. Thus, in the present preliminary study, performance was assessed according to both scoring criteria in order to ascertain whether the simple score would be as discriminative as the ceiling-floor score. The results indicated that that is actually the case. The comparison between the criteria was done by means of pairwise statistical analyses (between successive school levels as well as between successive age levels) following ANOVAs. Thus, for the present translation and collective application, the simple scoring was as discriminative as ceiling-floor scoring. In the sample, it was found a positive correlation between the children's school level and their parent's socio-economic level. The contribution of parent's socio-economic level was controlled via ANCOVAs. It was observed that the father's school level (but not the mother's) contributed to test performance. The standard error of measurement and internal consistency coefficients were obtained and revealed that the present translation of PPVT is surprisingly reliable, even before any item-reordering was done on the basis of difficulty

level shown by item analysis. Such reordering was subsequently done on the basis of item analysis performed on the data from the whole sample, and is offered in appendix. Finally, on the basis of both simple and ceiling-floor scores obtained from the collective application, normatization tables were generated, and are presented here as a function of both school level and age level. The present preliminary study aims at elaborating the definitive Brazilian adaptation of the PPVT.

Keywords: *Peabody Picture Vocabulary Test, Vocabulary, Psychological Testing*

O Teste de Vocabulário por Imagens Peabody: sua tradução, adaptação e computadorização

O vocabulário receptivo-auditivo e o Teste de Vocabulário por Imagens Peabody

O vocabulário receptivo-auditivo é um requisito para recepção e processamento de informação auditiva. Portanto, a extensão de vocabulário é uma importante medida de habilidade intelectual. De fato, tem sido demonstrado que a prova de vocabulário é o melhor índice individual de aproveitamento escolar (Dale & Reichert, 1957) e que as subprovas de vocabulário são as que mais contribuem para os escores das provas de inteligência geral (Elliott, 1983). O vocabulário está fortemente correlacionado com a habilidade de leitura e com o quociente de inteligência (Eysenck & Keane, 1990). A correlação com inteligência geral decorre do fato de que a aquisição do vocabulário requer o uso de informações contextuais para fazer inferências plausíveis sobre o significado de uma palavra desconhecida. Sternberg (1985) fornece um exemplo de como as inferências são empregadas para derivar significado de palavras desconhecidas a partir do contexto frasal em que aparecem: Numa sentença como "Ao amanhecer, o *blen* surgiu no horizonte brilhando intensamente" o significado (*sol*) da palavra desconhecida *blen* é dado pelo contexto frasal.

Este mesmo tipo de processamento ativo de construção de inferências que caracteriza a aquisição de vocabulário também caracteriza a compreensão de leitura.

Programas voltados à expansão do vocabulário tendem a ter forte impacto sobre a compreensão da leitura. Tem sido demonstrado (McKeon, Beck, Omanson & Pople, 1985) que boa parte do efeito de procedimentos voltados a aumentar a compreensão de leitura de leitores fracos por meio da expansão do vocabulário deriva do desenvolvimento de habilidades de inferências feitas em sentenças em que as palavras são usadas, mais do que simplesmente da rotineira aprendizagem de palavras juntamente com suas definições. Naquele estudo o método de ensino de vocabulário baseado em contexto produziu maior melhora na compreensão global de texto do que o método de ensino de definições. Leitores fracos mostraram maior capacidade de raciocinar com novas palavras quando eles eram ensinados a usar essas novas palavras em contextos diferentes e a unir palavras relacionadas do que quando eles simplesmente aprendiam as novas palavras e suas definições.

Já que programas voltados à expansão do vocabulário têm impactos tão relevantes, é essencial ser capaz de avaliar a sua eficácia relativa por meio de instrumentos estandardizados. O *Teste de Vocabulário por Imagens Peabody (TVIP)* avalia o desenvolvimento lexical no domínio receptivo, ou seja, as habilidades de compreensão de vocabulário de crianças entre 2a6m até 18a de idade. Ele fornece uma avaliação objetiva, rápida e precisa do vocabulário receptivo-auditivo em uma ampla variedade de áreas. As áreas incluem: pessoas; ações; qualidades; partes do corpo; tempo, natureza, lugares; objetos; animais; ferramentas e instrumentos; termos matemáticos. Desde sua versão original em inglês (Dunn, 1959), o teste tem sido revisado (Dunn & Dunn, 1981) e adaptado a outras línguas, como espanhol (Dunn, Padilla, Lugo, Dunn, 1986). *TVIP* correlaciona-se fortemente com a maior parte dos testes de vocabulário e de inteligência verbal. Uma

revisão da literatura sobre validade e fidedignidade da versão original em inglês pode ser encontrada em Robertson & Eisenberg (1981).

Descrição e indicações do *Teste de Vocabulário por Imagens Peabody*

O presente artigo baseia-se na adaptação hispano-americana do teste (Dunn e col., 1986). Nela, *TVIP* consiste em cinco pranchas de prática, seguidas de 125 pranchas de teste, organizadas em ordem crescente de dificuldade. As pranchas são compostas de quatro desenhos de linha preta em fundo branco. O teste é organizado de acordo com um modelo de múltipla escolha. Não requer que o examinando leia, escreva, ou vocalize qualquer coisa. A tarefa do examinando é selecionar a alternativa que melhor representa a palavra falada apresentada pelo examinador. Na presente tradução do teste, as 130 palavras faladas pelo examinador para avaliar o vocabulário receptivo-auditivo encontram-se listadas no Anexo 1. Na versão hispano-americana, o teste é administrado individualmente em cerca de 10 a 15 minutos. Para esta administração são adotados critérios de piso e teto, descritos mais abaixo.

Em termos de indicação, *TVIP* avalia o vocabulário receptivo-auditivo de examinandos entre 2a6m e 18a de idade. Pode ser empregado como prova de aproveitamento escolar, ou de inteligência verbal. É apropriado para avaliar o nível de desenvolvimento de linguagem receptiva em pré-escolares, bem como em crianças ou adultos que não lêem, que não escrevem e mesmo que não falam. Sua grande importância para a educação especial decorre do fato de que é especialmente apropriado para avaliar o vocabulário de pessoas incapazes de vocalizar palavras de modo inteligível, já que para avaliar a compreensão auditiva de palavras isoladas, o teste requer que o examinando apenas escolha a figura que corresponde à palavra falada pelo examinador. É ideal para a avaliação de crianças ou adultos com distúrbios emocionais, psicose, autismo, retardo mental, afasias

ou paralisia cerebral. É também ideal em estudos longitudinais, para avaliar o efeito de diversas variáveis sobre o desenvolvimento de vocabulário durante períodos extensos de tempo. Um de tais estudos longitudinais, por exemplo, avaliou os efeitos de generalização de ganhos de programas de tratamento de habilidades de linguagem em crianças com retardo no desenvolvimento da linguagem (e.g., Warren & Kaiser, 1986). No entanto, o teste é pouco recomendado para crianças com perda auditiva ou visual acentuada.

Objetivos do presente artigo

Apesar de ser usado há mais de 35 anos em todo o mundo como meio de avaliar o efeito de programas educacionais, o *TVIP* ainda é pouco conhecido no Brasil e não há em nosso país traduções, adaptações ou normas brasileiras publicadas para seu emprego. De fato, tanto quanto sabemos, no Brasil não há qualquer instrumento padronizado específico disponível para pesquisadores, educadores e clínicos para a avaliação de vocabulário, quer receptivo quer expressivo. No presente artigo foram feitas a tradução, a adaptação e a normatização do *TVIP* com o objetivo de buscar preencher esta lacuna, oferecendo assim uma versão brasileira do teste para uso por pesquisadores, educadores e clínicos. O artigo relata um estudo em que foi empregada a versão traduzida literalmente da adaptação hispano-americana do teste, que contém 130 itens e destina-se a aplicação individual. Dentre os frutos de tal estudo destacam-se tábuas de normatização dessa versão traduzida para respondentes entre seis e 14 anos; análises da confiabilidade do instrumento *TVIP*; análises do grau de dificuldade de cada uma das pranchas componentes do teste, etc. Tais análises do grau de dificuldade dos vários itens é que permitiram a reordenação das pranchas para a composição da adaptação brasileira. Tal reordenação é essencial para poder chegar a uma adaptação brasileira definitiva do teste a um novo idioma, e ocorreu, por exemplo, quando o

teste foi adaptado da versão original em inglês para a versão hispano-americana em espanhol por Dunn e col. (1986).

No presente estudo a versão traduzida foi aplicada coletivamente em sala de aula a uma amostra de 687 crianças entre pré-escola e oitava série, com idades variando de seis a 18 anos. Foram comparados dois critérios para avaliação das respostas: a pontuação piso-teto usual, própria para aplicação individual, e a pontuação simples, não-tradicional e mais apropriada para aplicação coletiva em sala de aula. A pontuação piso-teto era estabelecida de acordo com a idade e o desempenho da criança conforme critérios bastante complicados. Por outro lado, a pontuação simples consistia meramente no total de acertos, e foi proposta como apropriada para aplicação quando os itens do teste ainda não se encontram ordenados de acordo com o grau de dificuldade. Assim, nesse estudo o mesmo desempenho foi avaliado de acordo com os dois critérios para determinar se a pontuação simples seria tão discriminativa quanto a pontuação piso-teto. A contribuição do nível sócio-econômico foi controlada por meio de análise de covariância. Erro padrão de mensuração e coeficientes de consistência interna foram obtidos para avaliar se a presente tradução da adaptação hispano-americana do teste foi ou não confiável. Foi obtida a reordenação dos itens do teste com base no grau de dificuldade. Finalmente, a partir da aplicação coletiva, foram geradas tábuas de normatização para pontuações simples e piso-teto em função de nível escolar e de faixa etária.

O grande projeto e a versão computadorizada

O presente estudo é parte de um mega-projeto voltado ao desenvolvimento e adaptação de instrumentos para a avaliação, tratamento, educação, reeducação, comunicação, reabilitação e capacitação de pacientes com *déficits* sensoriais, motores e de processamento cognitivo. Nos últimos sete anos de desenvolvimento de tal mega-projeto,

foram computadorizados vários teste-padrão para a avaliação de uma série de habilidades básicas de entrada que são críticas ao sucesso de uma série de estratégias para o desenvolvimento de habilidades de leitura, escrita, comunicação, trabalho, etc. Exemplos desses são a *Escala de Maturidade Mental Columbia* (Burgemeister, Blum & Lorge, 1972) para a avaliação de raciocínio categorial; o subteste de arranjo de figuras das *Escalas de Inteligência Wechsler* para Adultos e para Crianças (Wechsler, 1991) para a avaliação de raciocínio sequencial; o *Teste Token* para Crianças (DeRenzi & Vignolo, 1962; DiSimoni, 1978) para paralisados cerebrais e para afásicos para avaliação da compreensão auditiva a habilidade para executar operações sob comando verbal; o *Teste de Maturidade de Leitura* (Campos, 1994) e o *Teste de Prontidão para a Leitura* (Kunz, 1979) para a avaliação de desempenho escolar em pré-alfabetizandos; o *Teste de Figuras Invertidas* ou *Reversal Test* (Edfeldt, 1955) para a identificação de dificuldades visuais eventualmente presente em dislexias disidéticas; e o próprio *Teste de Vocabulário por Imagens Peabody*, dentre outros. Além da computadorização desses testes-padrão, muitos outros testes computadorizados originais têm sido criados para a avaliação de habilidades imagéticas e fonológicas, tais como os testes de consciência fonológica em manipulação fonêmica e silábica (Capovilla, Colomi, Nico, Capovilla, 1995) para identificação de dificuldades de processamento em dislexias disfonológicas, e leitura (Capovilla, Macedo, Duduchi, Capovilla & Gonçalves, 1996), dentre outros.

Informações ulteriores sobre estes podem ser obtidas em Capovilla, 1993; 1994a, b; 1996; Capovilla & Macedo, 1994; Capovilla, Macedo, Duduchi, Raphael, Guedes, Capovilla & Gonçalves, 1996; Capovilla, Macedo, Duduchi, Soria & Capovilla, 1996; Capovilla, Macedo, Duduchi, Thiers, Capovilla, Gonçalves & 1995; Capovilla, Macedo, Gonçalves, Capovilla, Raphael & Duduchi, 1995; Capovilla, Macedo, Raphael, Capovilla, Gonçalves, Duduchi & Guedes, 1995; Capovilla, Macedo, Raphael, Duduchi, Moreira, Gonçalves &

Capovilla, 1995; Capovilla, Macedo, Seabra, Feitosa & Thiers, 1994; Feitosa, Macedo, Capovilla, Seabra & Thiers, 1994; Gonçalves, Macedo, Duduchi & Capovilla, 1995; Macedo, Capovilla, Gonçalves, Seabra, Thiers & Feitosa, 1994; Thiers, Capovilla, Macedo, Feitosa & Seabra, 1994).

A versão computadorizada para avaliação de pessoas com sérios distúrbios motores e da fala

Conforme mencionado acima, o teste teve uma versão computadorizada recentemente desenvolvida chamada *TVIP-Comp* (Capovilla, Guedes, Macedo, Thiers, Raphael, Capovilla, Duduchi, 1995; Capovilla, Macedo, Raphael, Capovilla, Gonçalves, Duduchi, Guedes, 1995; Guedes, Raphael, Capovilla, Macedo, Duduchi, 1995). Como a versão tradicional, objetiva avaliar o desenvolvimento lexical no domínio receptivo, ou seja, as habilidades de compreensão de vocabulário de pessoas incapazes de vocalizar palavras de modo inteligível ou mesmo de apontar. É executável em microcomputador AT 486 equipado com monitor SVGA e placa de voz digitalizada. Como periférico de *input* para registrar a resposta de examinandos com diversas e severas deficiências motoras, pode empregar uma tela sensível ao toque, ou um *mouse* comum dotado de pequena alavanca adaptada a um de seus botões, ou uma placa digitalizadora de voz, ou um detector de gemidos de fabricação caseira "hanzômetro" (Capovilla, Macedo, Feitosa, 1994), ou mesmo um detector de direção de olhar (Capovilla, 1994; Capovilla, Macedo, Duduchi, 1994). Quando a tela sensível ao toque não pode ser empregada, o programa faz varredura serial automática entre alternativas, em velocidade ajustável ao grau de dificuldade motora da criança. Neste caso, quando a alternativa desejada estiver sendo iluminada, a única resposta requerida do examinando é um movimento voluntário discreto de uma parte qualquer do corpo, ou a emissão de um som como um gemido ou grunhido, ou mesmo o olhar dirigido a

uma câmera. Foi demonstrado recentemente que as figuras do teste computadorizado têm clareza denotativa comparável àquela das figuras da versão tradicional (Raphael e cols., 1995). A disponibilidade de uma versão computadorizada torna ainda mais importante a realização de estudos de padronização do teste a partir de amostras brasileiras, que foi o objetivo do presente estudo.

Aplicações: individual na clínica versus coletiva na sala de aula; e pontuação piso-teto versus pontuação simples

Aplicação individual a partir de critérios de piso e teto, e seus problemas

A adoção de critérios de piso e teto foi feita para poupar examinandos de ter que ser submetidos a itens excessivamente fáceis ou difíceis. Para cálculo de piso, são computadas as oito respostas corretas consecutivas a partir de um item inicial que é determinado pela idade cronológica do examinando. Para cálculo de teto, o item final é determinado pela ocorrência de seis erros em oito respostas consecutivas. Uma das desvantagens da adoção dos critérios de piso e teto é limitar o teste a uma modalidade de aplicação individual. Uma desvantagem adicional é um forte aumento na demanda de trabalho braçal nas fases de aplicação e de avaliação, já que podem ocorrer várias complicações que dificultam o estabelecimento dos critérios de piso e teto, conforme explicado abaixo.

Complicações de critérios de piso e teto

Dunn e cols. (1986) ilustram um caso ideal e cinco complicações possíveis quando são adotados os critérios de piso e teto:

Caso ideal: sem complicações

No caso ideal, a partir do ponto de início (e.g., item 26), que é determinado de acordo com a idade cronológica do examinando (e.g., 6a10m), são aplicados os itens

sucessivos até que ocorra o primeiro erro (e.g., item 46). Caso tenham ocorrido pelo menos oito acertos consecutivos antes dele (e.g., itens 38-45), pode-se considerar o piso como tendo sido estabelecido. Deve-se então continuar rumo ao final (i.e., item 125) até que o examinando cometa seis erros em oito itens consecutivos (e.g., 61-68). A aplicação deve ser interrompida neste ponto (i.e., item final aplicado). Do número do item final aplicado (e.g., 68) deve ser subtraída a frequência de erros (e.g., 12). O resultado é o *score bruto*, aqui chamado de *pontuação piso-teto* (e.g., 56).

Complicação 1: piso retrógrado

Como primeiro exemplo de complicação, é possível que o examinando cometa um erro antes de conseguir completar oito acertos consecutivos a partir do ponto de início. Quando isto ocorre é necessário regredir desde o ponto de início, aplicando itens anteriores até obter oito acertos consecutivos. Por exemplo, um examinando de 5a11m começa pelo item 10. No item 13 ele comete um erro. Acerta nos itens 9, 8, 7, 6, e 5. Como acertou os itens de 5 a 12, pode-se considerar o piso como tendo sido estabelecido. Então, a testagem é retomada a partir do item 14. Além de um erro no item 13 e de um outro no item 17, comete seis outros erros entre os itens 20 a 28. Para a obtenção da pontuação piso-teto, são subtraídos os oito erros dos 28 itens. Isto resulta na pontuação piso-teto 20.

Complicação 2: dois ou mais pisos

Uma segunda complicação possível é a ocorrência de dois ou mais pisos. Isto se dá quando o nível de habilidade do examinando é superior à sua idade cronológica, e conseqüentemente o critério de piso levou a prova a começar com itens demasiadamente fáceis para ele. Quando isto ocorre, para obtenção da pontuação piso-teto, deve-se eliminar o(s) primeiro(s) piso(s) e computar apenas o último. Além disso, deve-se desconsiderar todos os erros anteriores ao último piso, já que teoricamente se supõe que todos os itens anteriores ao piso deveriam ser respondidos corretamente. Por exemplo, um examinando de

9a2m começa pelo item 60. Acerta todos os itens de 60 a 67, estabelecendo assim o primeiro piso de oito acertos consecutivos. Erra no item 68. Novamente, acerta todos de 69 a 76, o que estabelece um segundo piso de oito acertos consecutivos. Erra em 77, 78 e 80. Comete mais seis erros de 83 a 90. Como foram estabelecidos dois pisos, o primeiro é eliminado. Como o erro 68 é anterior ao segundo piso, é desconsiderado. A partir do item final, 90, são subtraídos apenas os nove erros cometidos a partir do segundo piso, resultando na pontuação piso-teto de 81.

Complicação 3: dois ou mais tetos

Uma terceira complicação possível é a ocorrência de dois ou mais tetos. Isto pode se dar quando o nível de habilidade do examinando for inferior à sua idade cronológica, e consequentemente o critério de piso levou a prova a começar com itens demasiadamente difíceis para ele. Isto é relativamente frequente quando se lida com examinandos com atraso de desenvolvimento. Quando isto ocorre, para obtenção a pontuação piso-teto, deve-se considerar apenas o primeiro teto e eliminar o(s) seguinte(s). Além disso, deve-se desconsiderar todos os itens posteriores ao primeiro teto, mesmo que sejam acertos, já que teoricamente se supõe que todos os itens posteriores ao primeiro teto deveriam ser respondidos incorretamente. Por exemplo, um examinando de 7a2m começa pelo item 38. Como ele comete um erro no item 40, portanto antes de estabelecer um piso, a testagem começa regressivamente a partir do item 37. A ocorrência frequente de erros impede a ocorrência de oito acertos consecutivos para o estabelecimento do piso, o que ocorre apenas entre os itens 2 e 9. Entre o ponto de início (item 38) e o piso (itens 2-9) ocorreram duas sequências que qualificam como tetos (seis erros em oito itens consecutivos), o primeiro 17-24, e o segundo 27-34. O segundo é eliminado. Portanto, para a obtenção a pontuação piso-teto são contados os erros cometidos até o item 24 (o último item do primeiro teto).

Complicação 4: ausência de piso

Uma quarta complicação possível é a ausência de piso. Isto pode ocorrer em examinandos jovens e imaturos, ou seja, com idade cronológica próxima ao extremo inferior e nível de habilidade inferior à idade cronológica. Quando isto ocorre, o piso é definido como o item 1, e o escore corresponde ao número de acertos até o teto. Por exemplo, um examinando de 5a6m começa pelo item 10. Como ele erra o primeiro item, a testagem começa regressivamente a partir do item 9. Como ele vai cometendo seis erros esporádicos, o piso não se estabelece, e a testagem regride até o item 1, que então se torna o piso automático. Como do item 3 ao 10 houve seis erros, o critério de teto foi estabelecido. Portanto, para a obtenção da pontuação piso-teto, são subtraídos os seis erros dos dez itens aplicados.

Complicação 5: ausência de teto

A última complicação arrolada pelos autores diz respeito à ausência de teto. Isto pode ocorrer em examinandos mais velhos e com desempenho superior, ou seja, com idade cronológica próxima ao extremo superior e nível de habilidade superior à idade cronológica. Quando isto ocorre, o teto é definido como o item 125, e a pontuação piso-teto corresponde a 125 menos o número de erros a partir do piso. Por exemplo, um examinando de 16a5m começa pelo item 90, e acerta todos os itens até 97, mas erra 98. O primeiro piso se estabelece de 90 a 97. Erra novamente 103, mas acerta de 104 a 111. O segundo piso se estabelece de 104 a 111. No entanto, entre 112 e 125 comete seis erros esparsos, e não suficientemente agrupados (i.e., seis erros em oito itens) para satisfazer o critério de teto. Portanto, o item 125 fica estabelecido automaticamente como teto. Para a obtenção da pontuação piso-teto são subtraídos de 125 apenas os seis erros desde o segundo piso (i.e., os dois erros anteriores ao primeiro piso são eliminados).

Argumentos em favor de estandardização com base em aplicação coletiva

De acordo com Dunn e col. (1986), a adoção de critérios de piso e teto é necessária porque "a administração de itens demasiadamente fáceis ou difíceis para um sujeito não teria nenhuma utilidade e provavelmente invalidaria os resultados" (p. 15). Em consequência, para permitir a adoção dos critérios de piso e teto, a aplicação deve ser restrita ao modo individual. Ou seja, restringir a coleta de dados à situação de aplicação individual é o preço que o pesquisador paga para poder adotar aqueles critérios. As complicações arroladas acima, no entanto, podem tornar o pesquisador menos disposto a pagar o preço, ou seja, a ter que restringir a aplicação do teste ao modo individual simplesmente para poder adotar aqueles critérios. Isto pode ocorrer especialmente com investigadores que estão mais interessados em pesquisa educacional do que clínica, em pesquisas sobre o efeito comparativo de diferentes métodos de ensino, e em pesquisas sobre desenvolvimento da linguagem que empregam modelos de grupo.

Pode-se argumentar também que, diferentemente de avaliações individuais ao estilo clínico, avaliações coletivas de desempenho em sala de aula não são estranhas às populações-alvo pré-escolar e escolar. Por conseguinte, como muitos outros testes, *TVIP* também poderia ser aplicado coletivamente. Além disso, o uso de tal aplicação coletiva parece também especialmente apropriado em países carentes de recursos. Parece-nos provável que, num país como o nosso, estandardizar um instrumento para aplicação individual quando ele poderia ser estandardizado para avaliação coletiva poderia resultar na prática na indesejável sub-utilização desse instrumento, devido a fatores como falta de verbas, de pessoal qualificado, de instrumentos, e de instalações apropriadas, além da superlotação das classes que força uma substituição da atenção individualizada pela coletiva. É pertinente lembrar, ainda, que delineamentos de grupo têm sido usados com frequência muitas vezes maior do que delineamentos de caso único em pesquisa escolar de efeitos de variáveis tais como métodos de ensino sobre o desenvolvimento de vocabulário, pesquisas

em que *TVIP* é especialmente apropriado. Portanto, fatores tais como adequação às características naturais de avaliação no ambiente escolar, adequação aos delineamentos de pesquisa, economia de tempo e de recursos humanos, etc., parecem fornecer argumentos convincentes em prol da estandardização do teste com base em aplicação coletiva em sala de aula.

No presente estudo a estandardização é feita com base em aplicação coletiva sem preocupação para com critérios de piso e teto. Já que não houve preocupação para com os critérios quando da aplicação do teste, seria ideal também não ter que se preocupar com eles quando da tabulação dos resultados. Prescindir daqueles critérios na fase de tabulação certamente reduz dramaticamente a demanda de tempo e de esforços. Como vantagem adicional desta medida, é provável que ela torne os resultados ainda mais representativos do desempenho dos respondentes na medida em que a tabulação passaria a incluir todo o desempenho do respondente na integridade dos itens aplicados, e não apenas naquilo que se concebe como sendo a "faixa crítica" entre piso e teto. Tal ganho em validade poderia ocorrer já que aquilo que o idealizador concebe como sendo a faixa crítica de desempenho de uma criança pode nem sempre corresponder exatamente à faixa crítica real. Isto pode ser ainda mais verdadeiro quando se trata de classes especiais.

No entanto, antes que a normatização para aplicação coletiva possa ser desenvolvida com base em escores simples, é necessário demonstrar pelo menos que a pontuação simples (ou porcentagem de acerto), obtida sem qualquer preocupação com critérios de piso e teto, é de fato equivalente à pontuação piso-teto, ou seja, ao escore calculado com base nos critérios estabelecidos de piso e teto. É este um dos objetivos do presente estudo.

Usando a tradução do *TVIP* para avaliar o desenvolvimento do vocabulário receptivo da pré-escola à oitava série (6a-14a): Normatização fluminense para aplicação coletiva

Arrazoado e objetivos

O presente estudo consiste numa adaptação brasileira do *Teste de Vocabulário por Imagens Peabody* para aplicação coletiva. Além da tradução dos itens do teste, o estudo oferece tábuas de pontuação normatizada feitas a partir de dados provenientes dessa aplicação coletiva. Tais dados consistem em pontuação simples (frequência de acertos nos 125 itens do teste) bem como em pontuação piso-teto tradicional (derivada dos complexos critérios descritos anteriormente). A normatização foi feita em termos de série escolar (de pré-escola a oitava série) e de idade (de seis a 14 anos). A normatização em termos de idade inclui também dados de desempenho de respondentes entre 15 e 18 anos ainda cursando a oitava série. Além disso, ao mesmo tempo, o estudo busca avaliar a presença de eventuais efeitos sistemáticos de uma série de variáveis de contexto dos respondentes. Além da faixa etária e do nível escolar que são inerentes à normatização, o presente estudo analisa os resultados em busca de eventuais efeitos de variáveis de contexto tais como o gênero e a dominância manual dos respondentes, o nível profissional e de escolaridade paterno e materno, etc.

Antes de elaborar as tábuas de pontuação normatizada para aplicação coletiva com base em pontuação simples, é preciso demonstrar que os resultados de uma mesma aplicação avaliados por pontuação simples são compatíveis com aqueles avaliados por meio do sistema tradicional de pontuação piso-teto. A partir de aplicação coletiva em toda a população escolar de uma escola pública de primeiro e segundo graus, o presente estudo compara o desempenho dos respondentes nas duas medidas de pontuação piso-teto e pontuação simples, de modo a avaliar o grau de correlação entre elas. De acordo com o arrazoado subjacente, caso as distribuições das duas pontuações sejam significativamente correlacionadas, e caso ambas as pontuações sejam afetadas semelhantemente pelas mesmas

variáveis, então a pontuação simples pode ser considerada como equivalente à pontuação piso-teto, o que permite prescindir de critérios piso-teto em aplicações coletivas no futuro.

De modo a descrever os dados em sua complexidade e permitir comparações posteriores, o presente estudo fornece normas baseadas em ambas as pontuações, simples e piso-teto. As primeiras podem ser preferidas em educação e pesquisa educacional baseada em delineamentos experimentais de grupo (e.g., Campbell & Stanley, 1966), enquanto que as últimas podem ser preferidas em clínica e pesquisa clínico-educacional baseada em delineamentos experimentais de caso único (e.g., Barlow & Hersen, 1984). Além disso, fornece normas baseadas tanto na faixa etária quanto na série escolar. Esta é mais uma vantagem da presente adaptação brasileira do teste em relação à adaptação hispano-americana. Naquela adaptação do teste foram fornecidas apenas normas baseadas em faixa etária, já que os sistemas educacionais porto-riquenho e mexicano diferem em termos de estruturação curricular das séries, o que não ocorre em nosso país.

Método

Participantes

Participaram 687 crianças fluminenses de pré-escola até a oitava série do Colégio de Aplicação da Universidade do Estado do Rio de Janeiro, com idades variando de 6a5m até 18a7m. A Figura 1 representa a distribuição de frequência de respondentes para cada uma das faixas etárias (à esquerda) e de níveis de escolaridade (à direita). Havia 357 meninos (52%) e 330 (48%) meninas. Dos participantes com lateralidade conhecida, 82,5% eram destros e 7,5% eram sinistros. Conforme representado na figura, o número de respondentes por série variava de 52 a 103. Havia 52 pré-escolares, 57 escolares na primeira série, 59 na segunda, 56 na terceira, 60 na quarta, 103 na quinta, 99 na sexta, 98 na sétima, e 103 na oitava. Já em termos de faixa etária, a distribuição de frequência era próxima à normal,

crecendo entre seis anos (26 respondentes) e 13 anos (107 respondentes), e decrescendo daí até 18 anos (apenas três respondentes). Isto sugere que, para a presente amostra, as normas baseadas em nível escolar devem ser mais confiáveis do que aquelas baseadas em faixa etária. A confiabilidade nesta deve estar restrita à faixa entre 7 e 15 anos, em que havia pelo menos 40 respondentes em cada um dos níveis de idade.

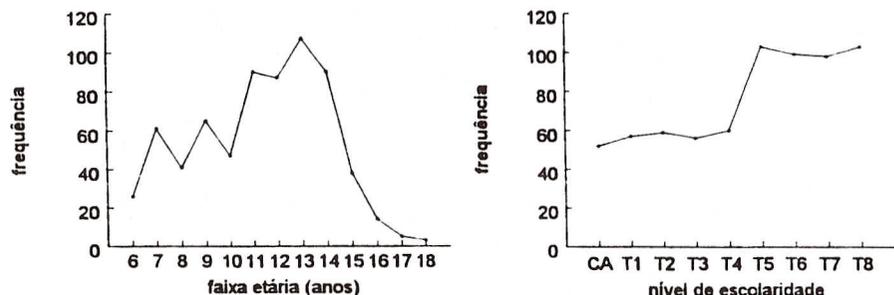


Figura 1. Distribuições de frequência dos respondentes em função de sua faixa etária (à esquerda) e de nível de escolaridade (à direita).

Os participantes eram provenientes de lares distribuídos em 99 bairros da Grande Rio, lares que eram representativos de cinco níveis econômico-culturais. Os níveis 1 a 5 variavam num crescente em termos de proventos e de nível educacional envolvido. De acordo com a convenção adotada no presente estudo, o nível 1 era constituído de pais desempregados ou aposentados e de mães aposentadas ou do lar; o nível 2 era constituído de pais empregados assalariados de nível baixo tais como operários, bancários, comerciários; o nível 3 era constituído de pais trabalhadores assalariados ou autônomos de nível médio, tais como técnicos, comerciantes, funcionários públicos e professores de primeiro e segundo graus; o nível 4 era constituído de pais profissionais liberais ou assalariados de nível superior, tais como médicos, dentistas, engenheiros, gerentes, professores de terceiro grau;

o nível 5 era constituído de pais empresários, industriais, e banqueiros. A escolaridade dos pais variava de 4 a 22 anos, com média de 13,6m. A Figura 2 representa as distribuições de frequência dos participantes em função do nível profissional (econômico-cultural) médio de seus pais na escala 1-5 (à esquerda), e do nível de escolaridade médio de seus pais em termos de categorias de anos cursados (à direita).

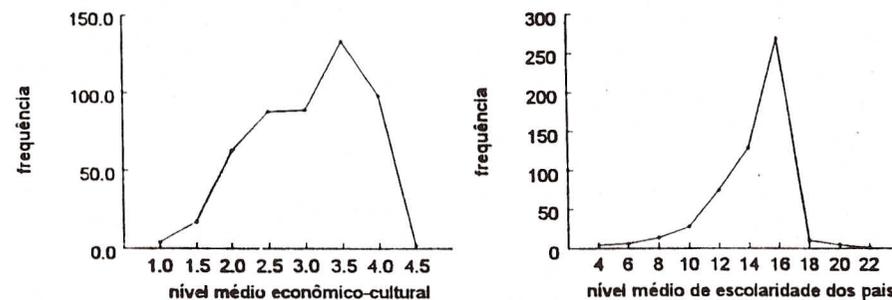


Figura 2. Distribuições de frequência dos respondentes da amostra em função do nível econômico-cultural médio dos pais (à esquerda), e do nível de escolaridade médio dos pais (à direita).

A distribuição de frequência de respondentes de cada nível econômico-cultural médio na amostra como um todo, bem como em cada nível de escolaridade encontra-se sumarizada na Tabela 1. Conforme sumarizado na tabela, à medida que aumentava o nível escolar dos respondentes, as distribuições do nível profissional médio dos pais tendiam a toma-se progressivamente mais negativas, com as medidas de tendência central progressivamente mais à direita da escala de escolaridade, especialmente entre terceira e oitava séries.

Tabela 1. Distribuições de frequência de respondentes, na amostra como um todo bem como em cada nível de escolaridade, em função do nível profissional médio de seus pais.

nível	total	CA	T1	T2	T3	T4	T5	T6	T7	T8
1,0	004	00	00	00	01	00	01	02	00	00
1,5	017	01	00	05	00	02	02	03	01	03
2,0	063	07	10	07	09	06	09	07	04	04
2,5	088	11	12	11	12	10	11	07	07	07
3,0	089	11	08	10	06	12	19	10	08	05
3,5	133	06	06	09	10	14	30	13	21	24
4,0	098	09	07	10	05	06	10	17	13	21
4,5	002	00	00	01	00	00	00	01	00	00
média	3,06	2,96	2,86	2,921	2,84	2,98	3,07	3,1	3,27	3,33
dp	0,75	0,71	0,70	0,82	0,73	0,68	0,68	0,88	0,65	0,72
md	3,0	3,0	2,5	3,0	2,5	3,0	3,0	3,5	3,5	3,5
inclin	-0,43	0,06	0,35	-0,10	-0,07	-0,32	-0,75	-0,65	-0,81	-1,09

Os dados acima sugerem a existência de correlação positiva entre o nível de escolaridade dos respondentes e o nível médio profissional dos pais. De fato, havia correlação positiva significativa entre o nível de escolaridade do respondente e o nível profissional médio dos pais (Pearson $r = 0,19$, $p < 0,0001$), bem como entre o nível de escolaridade dos pais e o nível de escolaridade dos respondentes (Pearson $r = 0,16$, $p < 0,0001$). Havia também correlação positiva significativa entre o nível de escolaridade e o nível econômico-cultural paterno (Pearson $r = 0,28$, $p < 0,0001$), mas não materno, bem como com o nível de escolaridade paterno (Pearson $r = 0,16$, $p < 0,003$), mas não materno. A Figura 3 representa a relação entre o nível profissional paterno e o nível de escolaridade

dos respondentes (à esquerda), e entre o nível de escolaridade paterno e o nível de escolaridade dos respondentes (à direita). Os respondentes de nível escolar mais avançado provinham de lares de nível profissional profissional e cultural mais elevado, provavelmente devido à evasão escolar, dentre outros fatores.

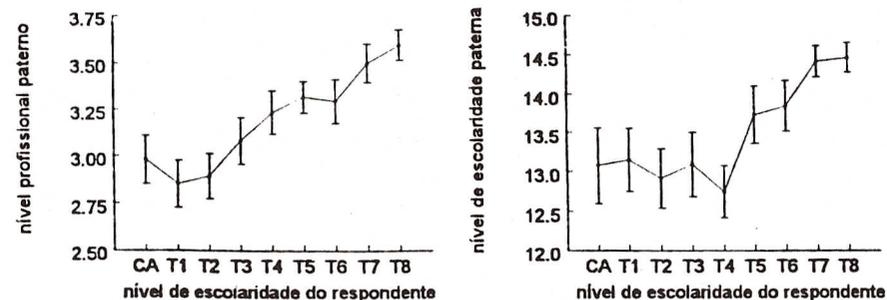


Figura 3. Relação entre escolaridade do respondente e nível profissional paterno (à esquerda) e nível escolaridade paterna (à direita). Dados representam média e erro padrão.

Uma vez que há estreita correlação entre faixa etária e escolaridade, como seria de se esperar, havia também correlação positiva significativa entre o faixa etária do respondente e o nível profissional médio dos pais (Pearson $r = 0,14$, $p = 0,002$), bem como entre o nível médio de escolaridade dos pais e a faixa etária dos respondentes (Pearson $r = 0,11$, $p = 0,012$). Também havia correlação positiva significativa entre a faixa etária dos respondentes e o nível profissional paterno paterno (Pearson $r = 0,21$, $p < 0,0001$), mas não materno, bem como com o nível de escolaridade paterno (Pearson $r = 0,14$, $p = 0,001$), mas não materno.

Aparato

O aparato consistia em lápis e uma versão do teste em formato reduzido. Esta consistia num pequeno caderno medindo 7 x 9 cm contendo 130 folhas com quatro figuras

cada uma. O lápis era empregado pelos respondentes para assinalar dentre as quatro figuras alternativas em cada folha do caderno aquela que correspondia à palavra falada pelo examinador. Cada respondente recebia seu próprio caderno individual que não era reutilizado após a aplicação. Portanto, para a presente pesquisa foram empregados 687 cadernos descartáveis. Foram também empregados uma tela branca e um retroprojeto com transparências correspondentes a cada uma das 130 folhas de resposta do caderno dos respondentes. O retroprojeto e as transparências eram empregados para assegurar que os respondentes estivessem com os cadernos abertos na página correta em que deveriam estar respondendo a qualquer dado momento, bem como para assegurar que as figuras do caderno fossem percebidas com suficiente nitidez e acuidade pelos respondentes. Deste modo, a cada item do teste os respondentes poderiam fazer sua escolha enquanto olhavam para a transparência projetada na tela, e usar o caderno meramente para assinalar a alternativa correta. A lista das cinco palavras de treino (marcadas de A a E) e das 125 palavras de teste (marcadas de 1 a 125) traduzidas para o português encontra-se no Anexo 2.

Procedimento

Antes do teste, logo de início as crianças eram informadas que sua participação era importante para ajudar os pesquisadores a desenvolver uma prova de vocabulário que por sua vez iria ser usada para ajudar as crianças com problemas escolares. A seguir cada criança recebia um caderno de 7 x 9 cm contendo 130 folhas com quatro figuras cada, em que devia assinalar com lápis a figura correspondente à palavra que era falada pelo examinador. As crianças eram informadas que sua tarefa era observar bem as quatro figuras de cada folha, e fazer um "X" sobre a figura que correspondia à palavra falada pelo examinador. Eram também informadas que para garantir que todos estivessem na folha

correta e que estivessem enxergando bem as figuras, o examinador iria projetar essas mesmas figuras na tela. Elas eram inicialmente expostas às cinco tentativas de treino e todas as perguntas eram respondidas conforme o procedimento-padrão de exame individual. Quando não havia mais questões, a testagem era iniciada a partir do item 1 para todas as crianças, independentemente da série escolar a que pertencessem.

As crianças eram examinadas coletivamente em sala de aula numa única sessão com duração variável, dependendo do ritmo de cada classe. Em cada um dos 130 itens do teste, o examinador projetava a transparência com as quatro figuras e falava em voz alta a palavra relativa à figura a ser escolhida. Repetia a palavra em voz alta após cinco segundos e concedia mais cinco a dez segundos para a escolha. Anunciava então que iria apresentar a próxima figura. Caso um dos respondentes da classe pedisse mais tempo, eram concedidos mais cinco a dez segundos. De modo geral, esperava-se que as crianças menores respondessem a cada item em cerca de 20 a 30 segundos, e as maiores em 10 a 20 segundos, sendo que o teste todo poderia ser aplicado no período de tempo aproximado à duração de uma aula.

Análise de dados

Dos registros escolares de cada respondente eram obtidos dados tais como idade em anos (6-18) e meses (77-223), gênero (masculino-feminino), dominância lateral (destro-sinistro), profissão paterna (110 tipos) e materna (80 tipos), nível profissional (1-5) e anos de escolaridade (4-22) paterna e materna, e finalmente bairro de residência (99 bairros). Tais informações eram dispostas numa planilha juntamente com os dados de desempenho: pontuação simples, pontuação piso-teto, duração da prova, e pontuação de acordo com as normatização hispânica.

Para cada respondente era tabulada a frequência de acertos nos 125 itens de teste. Tal frequência correspondia à pontuação simples. Em seguida era obtida a pontuação piso-teto conforme procedimento explicado anteriormente, e revisado a seguir em seu caso mais simples ("sem complicações"): Era buscado o piso que consistia nos últimos oito acertos consecutivos antes de um erro, desde o ponto de início que era estabelecido conforme a idade cronológica do respondente. Erros anteriores a este piso eram desconsiderados. Então eram buscados seis erros em oito itens consecutivos. O número do último item em que este critério era satisfeito era o teto. Deste número eram subtraídos os erros cometidos desde o piso. Este resto correspondia à pontuação piso-teto. Além da pontuação simples e da pontuação piso-teto, era registrada a duração da aplicação à turma a que pertencia o respondente. Finalmente, era também obtida a pontuação-hispânica de cada respondente de acordo com as normas-padrão hispânicas, de modo a permitir analisar sua correlação com as pontuações simples e piso-teto.

Para a estandardização da pontuação simples e da pontuação piso-teto, eram primeiramente obtidas a média e o desvio-padrão das distribuições dessas pontuações para cada um dos níveis escolares e para cada uma das faixas etárias. Para a obtenção da pontuação-padrão correspondente a cada pontuação de 1 a 125, tais pontuações eram submetidas à seguinte sequência de operações: de cada pontuação era subtraída a média da distribuição correspondente, e o resto era dividido pelo desvio-padrão dessa distribuição correspondente. Tal razão era multiplicada por 15, e a este produto era acrescido 100. A fórmula pode ser assim representada:

$$\text{pontuação-padrão} = ([\text{pontuação} - \text{média}]/\text{desvio padrão}) 15 + 100.$$

De modo a verificar o sentido e o grau de correlação entre a pontuação simples, a pontuação piso-teto, a pontuação de normas hispânicas, e a duração da aplicação em cada turma, coeficientes de correlação de Pearson eram calculados. Coeficientes de correlação

eram também calculados entre as variáveis independentes, de modo a identificar sinergias e antagonismos entre as variáveis, e identificar possíveis artefatos e contaminações de variáveis de interesse primário por outras de interesse secundário, e *vice-versa*.

ANOVAs foram feitas para analisar os efeitos das variáveis de interesse primário: faixa etária e nível de escolaridade, bem como das variáveis de interesse secundário: gênero, dominância lateral, nível profissional paterno e materno, anos de escolaridade paterno e materno. Quando variáveis primárias e secundárias encontravam-se significativamente correlacionadas, de modo a eliminar a eventual contribuição de uma variável na avaliação do efeito da outra variável, análises de covariância eram feitas. Por exemplo, uma variável secundária (e.g., nível profissional médio dos pais) poderia aparentar ter um grande efeito simplesmente porque ela se encontrava correlacionada a uma variável primária potente (nível escolar). A contribuição dessa variável primária era controlada ao torná-la covariante numa ANCOVA destinada a avaliar o efeito da secundária. Do mesmo modo, o efeito de uma variável primária poderia ser superestimado devido à contribuição de uma variável secundária com a qual encontrava-se correlacionada. De modo a isolar a contribuição da variável secundária à análise do efeito da primária, a secundária era tornada covariante numa ANCOVA destinada a avaliar o efeito da primária.

De modo a obter o coeficiente de consistência interna, a pontuação simples total de cada respondente era dividida em pontuação simples nos itens pares e pontuação simples nos itens ímpares. O coeficiente de correlação de Pearson era então calculado entre os itens pares e ímpares de todos os respondentes. Como tal procedimento usual reduz pela metade o comprimento do teste original, tal modo de calcular o coeficiente de consistência interna tende a subestimar a dimensão da verdadeira consistência interna. De modo a corrigir tal efeito de redução, a correção de Spearman-Brown (Downie & Heath, 1979) foi empregada. A fórmula de tal correção é assim sumarizada: $r_o = 2r_{pi}/(1+r_{pi})$, onde r_o representa a

confiabilidade (ou coeficiente de correlação) do teste original, r_{pi} , representa a confiabilidade (ou coeficiente de correlação) obtida pela correlação entre itens pares e ímpares. Assim, por exemplo, se um teste de 100 itens apresentasse um coeficiente de correlação par-ímpar de 0,85, a aplicação da correção Spearman-Brown corrigiria aquele coeficiente subestimado para 0,92.

Resultados

1. Comparação entre pontuação simples e pontuação piso-teto

Em termos de comparação entre a pontuação simples e a pontuação piso-teto, a Figura 4, à esquerda, representa a distribuição de frequência de classes de pontuação piso-teto e de pontuação simples na amostra como um todo. Dado o grande número de observações, as pontuações eram agrupadas em classes de três unidades cada uma, indo de 31-35 até 121-125. Conforme a figura, há uma razoável superposição entre as distribuições das pontuações obtidas pelos diferentes procedimentos de tabulação. As distribuições aparentam ter aproximadamente a mesma tendência central. De fato, os dados originais das distribuições de pontuações não-agrupadas mostram a mesma média (103,4), embora a distribuição da pontuação simples tenha dispersão inferior (desvio-padrão 15,0 contra 20,3) e mediana inferior (109 contra 112). A Figura 4, à direita, representa o correlograma entre a pontuação simples e a pontuação piso-teto, incluindo a reta de regressão e o intervalo de confiança. Conforme a figura, a correlação entre pontuações simples e piso-teto foi bastante elevada (coeficiente de correlação de Pearson $r = 0,967$, coeficiente de determinação $r^2 = 0,934$, $p < 0,0001$, $N = 680$). De fato a regressão da pontuação simples sobre a pontuação piso-teto resultou altamente significativa ($F_{(1,678)} = 9635,3$, $p < 0,0001$).

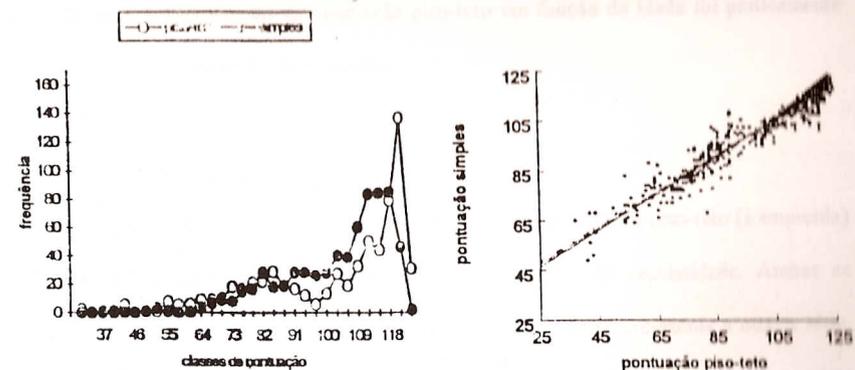


Figura 4. Distribuição de frequência de pontuação piso-teto e de pontuação simples na amostra como um todo (à esquerda), e correlograma entre pontuação simples e pontuação piso-teto (à direita), com reta de regressão e intervalo de confiança

2. Comparação dos efeitos das variáveis de interesse primário (nível de escolaridade e faixa etária dos respondentes) sobre pontuação simples e pontuação piso-teto

2.1 Efeitos de faixa etária

A Figura 5 representa a distribuição das médias da pontuação piso-teto (à esquerda) e de pontuação simples (à direita) em função da faixa etária. Conforme representado na figura, a pontuação piso-teto aumentou sistematicamente com uma aceleração praticamente constante entre seis e 11 anos de idade, sofreu então uma certa desaceleração do crescimento entre 11 e 14 anos, e a partir daí manteve-se assintoticamente estável até os 17-18 anos de idade. Os dados da pontuação simples são essencialmente semelhantes aos da pontuação piso-teto, guardadas as proporções, já que a amplitude de variação foi menor. Os dados mostram uma taxa elevada e estável de crescimento entre seis e 10-11 anos, uma redução na taxa de crescimento a partir daí com a manutenção de uma taxa moderada entre 10-11 e 14 anos, e uma parada no crescimento de 14 a 18 anos. Para ambas as distribuições,

com o aumento na faixa etária, além do aumento sistemático na média houve também uma redução sistemática na dispersão, indicando assim uma maior homogeneidade nos padrões de resposta. Tal redução na dispersão foi observada especialmente no período em que o crescimento se desacelerou entre 11 e 14 anos, mantendo-se a partir daí bastante baixa até os 17 anos.

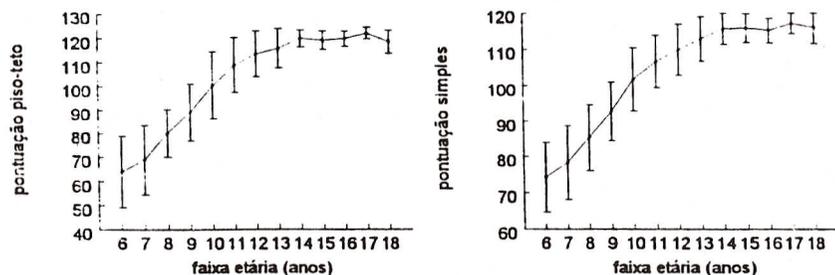


Figura 5. Média e desvio-padrão de pontuação piso-teto (à esquerda) e de pontuação simples (à direita) em função de faixa etária.

Antes de concluir que o teste não discrimina o crescimento de vocabulário após os 14 anos de idade, é preciso lembrar que a presente amostra era composta de respondentes apenas até a oitava série. Sendo assim, a faixa etária de 15 a 18 anos da presente amostra era composta essencialmente de retardatários, ou seja, de respondentes cujo desempenho estava abaixo do esperado para sua idade cronológica. Tais respondentes não eram representativos da população de estudantes dessa faixa etária, população esta que em sua maioria estaria matriculada no segundo grau. Assim, a constância observada na pontuação piso-teto entre 14 e 18 anos de idade pode significar simplesmente que o nível de vocabulário receptivo dos respondentes correspondia à série que cursavam. De fato, considerando apenas a faixa entre

seis e 14 anos, a distribuição da pontuação piso-teto em função de idade foi praticamente idêntica àquela em função de nível escolar.

2.2. Efeitos de nível de escolaridade

A Figura 6 representa a distribuição das médias da pontuação piso-teto (à esquerda) e da pontuação simples (à direita) como função do nível de escolaridade. Ambas as distribuições mostram um aumento sistemático nas médias entre pré-escola e oitava série. Ambas mostram uma taxa de crescimento elevada e relativamente constante até a quinta série, quando então aparentaram diminuir. Ao lado da desaceleração da pontuação, nesta mesma faixa da quarta à oitava série, chama a atenção a redução na dispersão dos dados. A desaceleração no crescimento das médias e a redução na dispersão parecem ser indicativas de um efeito de teto no teste. De fato, da sétima à oitava série houve apenas uma pequena redução na dispersão, sendo que a média da pontuação simples não cresceu e a da pontuação piso-teto cresceu praticamente nada. Isto se coaduna com a hipótese de efeito de teto. No entanto tal hipótese não parece se coadunar com os dados de Dunn e col. (1986) para amostras porto-riquenha e mexicana. Naquele estudo a pontuação cresce de 13 a 14 e continua a crescer entre 14 e 18 anos, se bem que com menor taxa. Para a população brasileira, no entanto, a hipótese de teto só pode ser refutada com clareza a partir da coleta de dados com escolares de segundo grau.

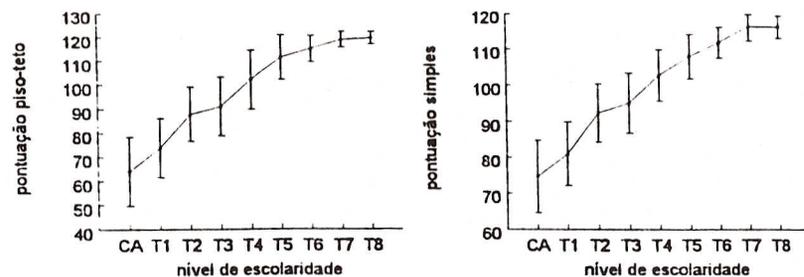


Figura 6. Média e desvio-padrão de pontuação piso-teto (à esquerda) e de pontuação simples (à direita) em função de nível de escolaridade.

3. Análises estatísticas

A grande similaridade nas distribuições de pontuação simples e de pontuação piso-teto tanto em função de faixa etária quanto em função de nível de escolaridade sugere fortemente que essas medidas sejam equivalentes, guardadas as proporções, já que a amplitude da variação na distribuição das médias da pontuação simples foi menor. De maneira a testar estatisticamente esta aparência de semelhança, ANOVAs foram conduzidas, e testes de significância da diferença entre pares foram empreendidos, tanto para faixa etária quanto para nível de escolaridade. O objetivo era verificar se o efeito de cada uma dessas variáveis tomada isoladamente diferia ou não de maneira sutil nas duas pontuações.

3.1. Efeito do nível de escolaridade (série escolar) do respondente

3.1.1. Sobre pontuação piso-teto

ANOVA revelou efeito significativo do nível de escolaridade dos respondentes sobre sua pontuação piso-teto ($F_{[8, 671]} = 333,81, p < 0,0001$). A pontuação simples foi de 64,1 na pré-escola, 73,9 na primeira série, 87,9 na segunda, 91,1 na terceira, 102,3 na quarta, 111,7

na quinta, 115,4 na sexta, 119,4 na sétima, e 120,1 na oitava. Análises da significância das diferenças entre os pares via teste de Bonferroni revelaram que a pontuação piso-teto cresceu significativamente ($p < 0,0001$) de uma série à série seguinte apenas de pré-escola à primeira série, de primeira à segunda série, de terceira à quarta série, e de quarta à quinta série. As diferenças entre segunda e terceira séries, bem como entre quinta e sexta, sexta e sétima, e sétima e oitava séries não foram significantes ($p > 0,05$).

3.1.2. Sobre pontuação simples

ANOVA também revelou efeito significativo do nível de escolaridade dos respondentes sobre sua pontuação simples ($F_{[8, 671]} = 372,78, p < 0,0001$). A pontuação simples foi de 74,6 na pré-escola, 81 na primeira série, 92,3 na segunda, 91,1 na terceira, 102,8 na quarta, 108,1 na quinta, 112 na sexta, 116,4 na sétima, e 116,3 na oitava. Análises da significância das diferenças entre os pares via teste de Bonferroni revelaram que a pontuação simples cresceu significativamente ($p < 0,0001$) de uma série à série seguinte em todas as séries, exceto da segunda à terceira série, e da sétima à oitava série.

3.1.3. Conclusão

Assim, o método de tabulação por meio de pontuação simples parece ter sido mais discriminativo dos efeitos do nível de escolaridade sobre o desempenho nesse teste em vocabulário do que o método de tabulação de pontuação que levava em consideração os critérios de piso e teto. Considerando as oito diferenças de pontuação entre os nove níveis da pré-escola à oitava série, enquanto seis das oito diferenças em pontuação simples foram significantes, apenas quatro das oito diferenças em pontuação piso-teto atingiram significância. Isto sugere que a pontuação simples pode ser mais sensível às diferenças em desempenho no teste do que a pontuação piso-teto.

3.2. Efeito da faixa etária do respondente

3.2.1. Sobre pontuação piso-teto

ANOVA revelou efeito significativo do nível de escolaridade dos respondentes sobre sua pontuação piso-teto ($F_{(12,655)} = 170,22, p < 0,0001$). A pontuação piso-teto foi de 64,0 aos seis anos, 69,0 aos sete, 80,1 aos oito, 88,9 aos nove, 100,3 aos dez, 108,7 aos onze, 113,3 aos 12 anos, 115,5 aos 13 anos, 119,6 aos 14 anos, 118,8 aos 15 anos, 119,6 aos 16 anos, 121,8 aos 17 anos, e 118,3 aos 18 anos. Análises da significância das diferenças entre os pares via teste de Bonferroni revelaram que a pontuação piso-teto não cresceu significativamente ($p > 0,05$) entre 6 e 7 anos, mas cresceu significativamente ($p < 0,0001$) entre sete e oito, entre oito e nove, entre nove e dez, e entre dez e onze anos. A partir daí, as pontuações entre um ano e o seguinte não mais cresceram significativamente ($p > 0,05$). Assim, não houve crescimento significativo nas pontuações entre 11 e 12, entre 12 e 13, entre 13 e 14, entre 14 e 15, entre 15 e 16, entre 16 e 17, ou entre 17 e 18 anos. Embora a pontuação aos 11 anos não fosse significativamente menor que aquela aos 12 anos, ela foi significativamente menor que aquelas aos 13, 14, 15, e 16 anos. Além disso, houve crescimento significativo na pontuação entre 12 e 14 anos. Não houve qualquer diferença significativa entre as pontuações piso-teto entre quaisquer anos entre 13 e 18 anos.

3.2.2. Sobre pontuação simples

ANOVA revelou efeito significativo do nível de escolaridade dos respondentes sobre sua pontuação simples ($F_{(12,655)} = 181,26, p < 0,0001$). A pontuação simples foi de 74,4 aos seis anos, 78,4 aos sete, 85,4 aos oito, 92,8 aos nove, 101,7 aos dez, 106,7 aos onze, 110,0 aos 12 anos, 113,0 aos 13 anos, 115,7 aos 14 anos, 115,9 aos 15 anos, 115,2 aos 16 anos, 117,2 aos 17 anos, e 116,0 aos 18 anos. Análises da significância das diferenças entre os

pares via teste de Bonferroni revelaram que a pontuação piso-teto não cresceu significativamente ($p > 0,05$) entre 6 e 7 anos, mas cresceu significativamente ($p < 0,001$) entre sete e oito, entre oito e nove, entre nove e dez, e entre dez e onze anos. A partir daí, as pontuações entre um ano e o seguinte não mais cresceram significativamente ($p > 0,05$). Assim, não houve crescimento significativo nas pontuações entre 11 e 12, entre 12 e 13, entre 13 e 14, entre 14 e 15, entre 15 e 16, entre 16 e 17, ou entre 17 e 18 anos. Embora a pontuação aos 11 anos não fosse significativamente menor que aquela aos 12 anos, ela foi significativamente menor que aquelas aos 13, 14, 15, e 16 anos. Além disso, houve crescimento significativo na pontuação entre 12 e 14 anos, bem como entre 12 e 15 anos. Não houve qualquer diferença significativa entre as pontuações piso-teto entre quaisquer anos entre 13 e 18 anos.

3.2.3. Conclusão

Assim, o método de tabulação por meio de pontuação simples parece ter sido rigorosamente tão discriminativo quanto o método de tabulação piso-teto para a identificação dos efeitos do nível de escolaridade sobre o desempenho nesse teste de vocabulário. Considerando as doze diferenças de pontuação entre as 13 idades da faixa, as idades em que houve diferença na pontuação simples foram as mesmas idades em que houve diferença na pontuação piso-teto: entre 7 e 8, entre 8 e 9, entre 9 e 10, e entre 10 e 11 anos. Houve correspondência ponto a ponto de significância em todas as diferenças entre as pontuações derivadas dos dois métodos de tabulação, com uma pequena diferença: a pontuação simples discriminou uma diferença de desempenho entre 12 e 15 anos que a pontuação piso-teto não havia identificado. Em conclusão, pode-se dizer que a distribuição de pontuação simples por faixa etária foi rigorosamente equivalente à distribuição da pontuação piso-teto. Quando duas variáveis são afetadas em precisamente a mesma maneira no mesmo grau por uma

terceira variável, pode-se dizer que elas são equivalentes entre si em termos dessa terceira variável. Assim, no presente estudo, em termos de faixa etária, foi observada equivalência entre pontuação simples e pontuação piso-teto.

3.3. Regressão das pontuações em faixa etária

A Figura 7, à esquerda, representa a regressão da pontuação piso-teto sobre a faixa etária, em meses. Conforme representado na figura, a pontuação piso-teto esteve fortemente correlacionada com a faixa etária em meses (coeficiente de correlação de Pearson $r = 0,83$, $N = 667$), sendo que 68 por cento da variância na pontuação piso-teto foram explicados pela variação na faixa etária ($r^2 = 0,68$). A regressão da pontuação piso-teto sobre a faixa etária em meses resultou significativa ($F_{[1,665]} = 1439,8$, $p < 0,0001$). A regressão da mesma pontuação piso-teto sobre a faixa etária em anos também foi significativa ($F_{[12,655]} = 170,22$, $p < 0,0001$), sendo a correlação ainda maior ($r = 0,87$, $r^2 = 0,76$, $N = 668$). A Figura 7, à direita, representa a regressão da pontuação simples sobre a faixa etária. Conforme a figura, a pontuação simples esteve ainda mais fortemente correlacionada com a faixa etária do que a pontuação piso-teto (Pearson $r = 0,84$, $N = 667$), sendo que 70 por cento da variância na pontuação simples foram explicados pela variação na faixa etária ($r^2 = 0,70$). A regressão da pontuação simples sobre a faixa etária também resultou significativa, sendo que o tamanho do efeito foi ainda maior ($F_{[1,665]} = 1581,0$, $p < 0,0001$). A regressão da mesma pontuação simples sobre a faixa etária em anos também foi significativa ($F_{[12,655]} = 181,26$, $p < 0,0001$), sendo a correlação ainda maior ($r = 0,88$, $r^2 = 0,77$, $N = 668$).

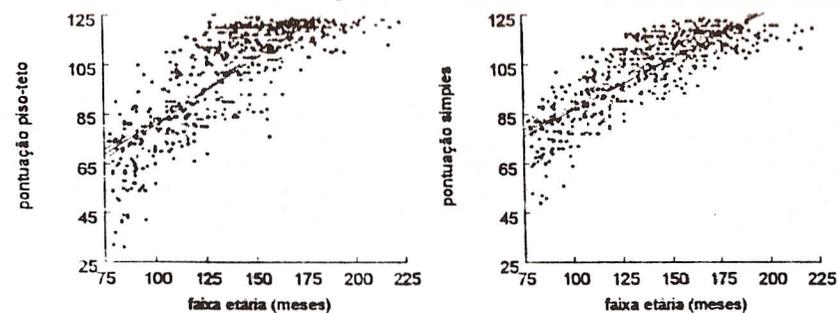


Figura 7: Regressão de pontuação piso-teto (à esquerda), e de pontuação (à direita) sobre faixa etária.

Também resultaram significantes a regressão da pontuação piso-teto sobre o nível de escolaridade ($r = 0,87$, $r^2 = 0,75$, $N = 680$, $F_{[1,678]} = 2055,8$, $p < 0,0001$), bem como a regressão da pontuação simples sobre o nível de escolaridade ($r = 0,88$, $r^2 = 0,78$, $N = 680$, $F_{[1,678]} = 2373,2$, $p < 0,0001$).

3.4. Análises de covariância: controlando o efeito dos níveis profissional e de escolaridade paternos, e dos níveis médios profissional e de escolaridade dos pais.

Conforme descrito na seção sobre os participantes do presente estudo, havia uma correlação significativa entre, de um lado, o nível de escolaridade dos respondentes e, de outro lado, os níveis profissional e de escolaridade paterno, e os níveis médios profissional e de escolaridade dos pais. É possível que tais correlações positivas possam contribuir para inflacionar de algum modo a estimativa acerca dos efeitos do nível de escolaridade dos respondentes sobre a pontuação, que haviam sido identificados por meio da ANOVA. Ou seja, é possível que uma parte do efeito do nível de escolaridade sobre a pontuação seja relacionado indiretamente àquelas variáveis de fundo. Por meio de ANCOVAs os efeitos

daqueles quatro covariantes foram isolados de modo que a estimativa do efeito do nível de escolaridade do respondente sobre as pontuações obtidas pudesse ser analisado. As ANCOVAs aqui realizadas controlaram os efeitos daqueles quatro covariantes simultaneamente. As ANCOVAs revelaram significância do efeito do nível de escolaridade dos respondentes tanto sobre a pontuação piso-teto ($r = 0,88$, $r^2 = 0,78$, $N = 474$, $F_{[8,461]} = 188,2$, $p < 0,000$) quanto sobre a pontuação simples ($r = 0,9$, $r^2 = 0,81$, $N = 474$, $F_{[8,461]} = 220,5$, $p < 0,0001$). As mesmas ANCOVAs foram feitas para analisar o efeito de faixa etária sobre ambos os escores controlando o efeito dos quatro covariantes simultaneamente. Também essas ANCOVAs revelaram significância do efeito da faixa etária dos respondentes tanto sobre a pontuação piso-teto ($r = 0,87$, $r^2 = 0,75$, $N = 463$, $F_{[12,446]} = 101,38$, $p < 0,0001$) quanto sobre pontuação simples ($r = 0,88$, $r^2 = 0,78$, $N = 463$, $F_{[12,446]} = 116,56$, $p < 0,0001$). Tais dados sugerem que ao avaliar o vocabulário receptivo dos respondentes, o teste é sensível aos efeitos de faixa etária e de escolaridade, sendo suficientemente robusto a artefatos e efeitos de covariantes espúrios.

4. Comparação dos efeitos das variáveis de interesse secundário (níveis profissionais paterno, materno e médio, níveis de escolaridade paterno, materno e médio) sobre pontuação simples e pontuação piso-teto

4.1. O efeito do nível profissional e de escolaridade dos pais

Conforme mencionado na descrição dos participantes, havia correlação positiva significativa do nível de escolaridade com o nível econômico-cultural paterno ($r = 0,28$, $p < 0,0001$), mas não materno, bem como com o nível de escolaridade paterno ($r = 0,16$, $p < 0,003$), mas não materno. A existência daquelas correlações positivas significantes com o nível de escolaridade do respondente poderia fazer com que o nível médio econômico e cultural dos pais, o nível econômico-cultural paterno, e o nível de escolaridade paterno

aparentassem ter efeito significativo sobre o desempenho no teste. Se tal aparência fosse falsa, aquele efeito aparente seria um artefato de correlação espúria com a variável verdadeiramente efetiva, nível de escolaridade do respondente. De modo a verificar se as três variáveis aqui arroladas (nível econômico-cultural médio e paterno, e nível de escolaridade paterno) têm efeito sobre o desempenho no teste é necessário subtrair o efeito da variável correlacionada (nível de escolaridade do respondente) por meio de uma análise de covariância. Quanto maior a correlação entre a variável de interesse (e.g., nível econômico-cultural ou nível de escolaridade) e a variável a ela correlacionada (e.g., nível de escolaridade do respondente), tanto maior é a redução na variância de erro obtida por meio da análise de covariância (ANCOVA, Kirk, 1982).

Quando as três variáveis correlacionadas significativamente com nível de escolaridade do respondente têm subtraído o efeito desse covariante via ANCOVA, seu efeito sobre o desempenho no teste resulta não-significante. Por outro lado, os níveis maternos econômico-cultural e de escolaridade não estavam correlacionados significativamente ao nível de escolaridade dos respondentes. Quando seus efeitos sobre o desempenho dos respondentes no teste são analisados via ANCOVA, o nível de escolaridade materno mostrou-se significativo ($F_{[7,539]} = 2,06$, $p = 0,046$), mas apenas para pontuação simples, não para pontuação piso-teto. Não houve evidência de qualquer efeito de nível profissional e educacional dos pais sobre a pontuação piso-teto

4.2. Conclusão

A similaridade das distribuições de pontuação simples e pontuação piso-teto, a elevada e significativa correlação entre elas, o fato de que elas são afetadas pelas mesmas variáveis em aproximadamente a mesma medida, o fato de que elas igualmente não parecem ser afetadas por outras variáveis, são todas evidências que sugerem fortemente que a

daqueles quatro covariantes foram isolados de modo que a estimativa do efeito do nível de escolaridade do respondente sobre as pontuações obtidas pudesse ser analisado. As ANCOVAs aqui realizadas controlaram os efeitos daqueles quatro covariantes simultaneamente. As ANCOVAs revelaram significância do efeito do nível de escolaridade dos respondentes tanto sobre a pontuação piso-teto ($r = 0,88$, $r^2 = 0,78$, $N = 474$, $F_{[8,461]} = 188,2$, $p < 0,000$) quanto sobre a pontuação simples ($r = 0,9$, $r^2 = 0,81$, $N = 474$, $F_{[8,461]} = 220,5$, $p < 0,0001$). As mesmas ANCOVAs foram feitas para analisar o efeito de faixa etária sobre ambos os escores controlando o efeito dos quatro covariantes simultaneamente. Também essas ANCOVAs revelaram significância do efeito da faixa etária dos respondentes tanto sobre a pontuação piso-teto ($r = 0,87$, $r^2 = 0,75$, $N = 463$, $F_{[12,446]} = 101,38$, $p < 0,0001$) quanto sobre pontuação simples ($r = 0,88$, $r^2 = 0,78$, $N = 463$, $F_{[12,446]} = 116,56$, $p < 0,0001$). Tais dados sugerem que ao avaliar o vocabulário receptivo dos respondentes, o teste é sensível aos efeitos de faixa etária e de escolaridade, sendo suficientemente robusto a artefatos e efeitos de covariantes espúrios.

4. Comparação dos efeitos das variáveis de interesse secundário (níveis profissionais paterno, materno e médio, níveis de escolaridade paterno, materno e médio) sobre pontuação simples e pontuação piso-teto

4.1. O efeito do nível profissional e de escolaridade dos pais

Conforme mencionado na descrição dos participantes, havia correlação positiva significativa do nível de escolaridade com o nível econômico-cultural paterno ($r = 0,28$, $p < 0,0001$), mas não materno, bem como com o nível de escolaridade paterno ($r = 0,16$, $p < 0,003$), mas não materno. A existência daquelas correlações positivas significantes com o nível de escolaridade do respondente poderia fazer com que o nível médio econômico e cultural dos pais, o nível econômico-cultural paterno, e o nível de escolaridade paterno

aparentassem ter efeito significativo sobre o desempenho no teste. Se tal aparência fosse falsa, aquele efeito aparente seria um artefato de correlação espúria com a variável verdadeiramente efetiva, nível de escolaridade do respondente. De modo a verificar se as três variáveis aqui arroladas (nível econômico-cultural médio e paterno, e nível de escolaridade paterno) têm efeito sobre o desempenho no teste é necessário subtrair o efeito da variável correlacionada (nível de escolaridade do respondente) por meio de uma análise de covariância. Quanto maior a correlação entre a variável de interesse (e.g., nível econômico-cultural ou nível de escolaridade) e a variável a ela correlacionada (e.g., nível de escolaridade do respondente), tanto maior é a redução na variância de erro obtida por meio da análise de covariância (ANCOVA, Kirk, 1982).

Quando as três variáveis correlacionadas significativamente com nível de escolaridade do respondente têm subtraído o efeito desse covariante via ANCOVA, seu efeito sobre o desempenho no teste resulta não-significante. Por outro lado, os níveis maternos econômico-cultural e de escolaridade não estavam correlacionados significativamente ao nível de escolaridade dos respondentes. Quando seus efeitos sobre o desempenho dos respondentes no teste são analisados via ANCOVA, o nível de escolaridade materno mostrou-se significativo ($F_{[7,539]} = 2,06$, $p = 0,046$), mas apenas para pontuação simples, não para pontuação piso-teto. Não houve evidência de qualquer efeito de nível profissional e educacional dos pais sobre a pontuação piso-teto

4.2. Conclusão

A similaridade das distribuições de pontuação simples e pontuação piso-teto, a elevada e significativa correlação entre elas, o fato de que elas são afetadas pelas mesmas variáveis em aproximadamente a mesma medida, o fato de que elas igualmente não parecem ser afetadas por outras variáveis, são todas evidências que sugerem fortemente que a

pontuação simples pode ser tomada como equivalente à pontuação piso-teto em aplicação coletiva. Tais evidências contribuem assim para legitimar a estratégia de normatização com base em pontuação simples.

5. Tempo dispendido em responder ao teste: Efeito de nível de escolaridade e de faixa etária, e correlação com pontuações simples e piso-teto

A Figura 8 representa a média e o desvio-padrão do tempo dispendido (min) em responder ao teste em função da faixa etária (à esquerda) e do nível de escolaridade (à direita). Como pode ser observado, a duração das sessões foi função inversa da faixa etária e do nível de escolaridade, variando de 65 min para os grupos menos maduros até 35 min para os mais maduros. Como as figuras revelam, a curva de tempo em função de faixa etária foi mais regular do que a curva de tempo em função de nível de escolaridade. Isto pode ter ocorrido em razão de variáveis estranhas ao teste e inerentes à situação de aplicação, já que a aplicação coletiva do teste ocorria em função de diferentes níveis de escolaridade, permitindo assim que variáveis estranhas operando naquela situação pudessem ter efeito generalizado nesse nível.

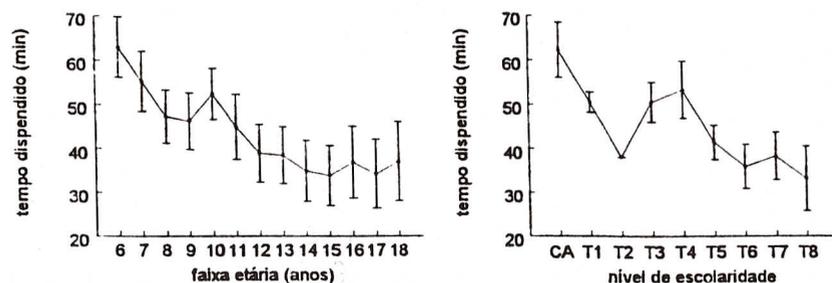


Figura 8. Média e desvio-padrão de tempo dispendido (min) na turma em função de faixa etária (à esquerda) e de nível de escolaridade (à direita).

Os dados foram obtidos a partir de aplicação coletiva em sala de aula. Escore e pontuação foram obtidos para idades entre 6 e 18 anos distribuídos de 6 em 6 meses, e para os nove níveis de escolaridade. Num primeiro momento foram comparados a pontuação simples e o escore obtido a partir de critérios de piso e teto. Regressões de escore e de pontuação nos 24 grupos etários resultaram significantes ($F_{[23,286]} = 29,31, p < 0,000$ e $F_{[23,286]} = 25,22, p < 0,000$, respectivamente), bem como nos nove níveis de escolaridade ($F_{[8,310]} = 108,25, p < 0,000$ e $F_{[8,310]} = 91,22, p < 0,000$, respectivamente). A regressão de escore sobre pontuação foi significativa ($r = 0,91, r^2 = 0,82, F_{[1,317]} = 1428,04, p < 0,000$). O fato de que a pontuação simples equivale ao escore obtido a partir da aplicação daqueles critérios sugere que os complexos critérios de piso e teto calculados a partir do desempenho individual são prescindíveis. Isto aumenta a confiança na adequação da estratégia de aplicação coletiva do teste, abrindo assim caminho à normatização a partir desta estratégia.

Houve correlação positiva significativa entre pontuação simples e pontuação piso-teto ($r = 0,97, p < 0,0001, N = 680$), conforme representado na Figura 7 (à esquerda); bem como entre de um lado pontuação padrão conforme normas hispano-americanas e de outro lado pontuação simples ($r = 0,72, p < 0,0001, N = 662$), conforme representado na Figura 7 (à direita), e pontuação piso-teto ($r = 0,78, p < 0,0001, N = 662$), como representado na Figura 8 (à esquerda). Houve correlação negativa significativa entre de um lado o tempo dispendido em responder ao teste na turma e de outro lado a pontuação simples ($r = -0,64, p < 0,0001, N = 617$) (Figura 8, à direita), a pontuação piso-teto ($r = -0,65, p < 0,0001, N = 617$) (Figura 9, à esquerda), e a pontuação padrão conforme normas hispano-americanas ($r = -0,33, p < 0,0001, N = 601$) (Figura 9, à direita).

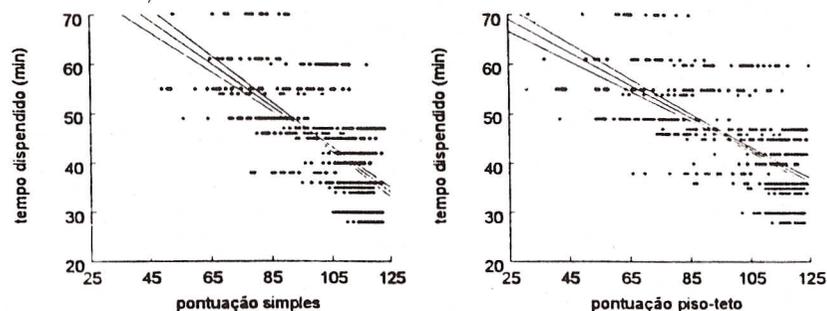


Figura 9. Correlograma entre tempo dispendido e pontuação simples (à esquerda), e entre tempo dispendido e pontuação piso-teto (à direita). Reta de regressão e intervalo de confiança.

A confiabilidade de PPVT na presente tradução

Foi calculado o coeficiente de consistência interna pelo método de divisão ao meio (*split-half*) (Downie & Heath, 1979). Diferentemente de outros métodos para obtenção de coeficientes de confiabilidade tais como o método de teste e reteste que requer a administração do mesmo teste mais de uma vez, e do método de formas paralelas que requer a administração de formas equivalentes do mesmo teste, o método de divisão ao meio requer apenas uma administração de um dado teste. Neste método, a pontuação de cada respondente é dividida em duas pontuações distintas, uma para os itens pares e outra para os itens ímpares do teste. O coeficiente de correlação de Pearson é então calculado entre as duas pontuações (i.e., pontuação 1 nos itens pares, e pontuação 2 nos itens ímpares) para toda a amostra de respondentes. No presente estudo a pontuação simples de cada respondente foi dividida em duas: a pontuação simples obtida nos itens pares e a pontuação simples obtida nos itens ímpares. O cálculo do coeficiente de correlação de Pearson entre

pontuação simples obtida nos itens pares e aquela obtida nos itens ímpares para todos os 680 respondentes foi de 0,914, o que atinge o nível de significância $p < 0,0001$.

Considerando que a confiabilidade de um teste está diretamente relacionada ao comprimento deste teste, quando o coeficiente de consistência interna é calculado por meio do método de divisão ao meio, o comprimento do teste original é em verdade cortado pela metade. Consequentemente, a estimativa de consistência obtida por meio desse método de divisão ao meio tende a ser artificialmente menor do que o coeficiente é de fato. Para corrigir tal distorção, a fórmula de correção de Spearman-Brown, apresentada abaixo, tem sido recomendada (Downie & Heath, 1979): $r_{ii} = 2r_{pi} / (1 + r_{pi})$,

em que r_{ii} corresponde ao coeficiente de confiabilidade do teste original, e r_{pi} corresponde ao coeficiente de confiabilidade obtido pelo método de divisão ao meio par-ímpar. Considerando que no presente estudo o coeficiente de 0,914 foi obtido pelo método de divisão ao meio, a aplicação da fórmula de correção de Spearman-Brown resulta no coeficiente de confiabilidade $r_{ii} = 0,955$.

O coeficiente de consistência interna foi também obtido pela fórmula 20 de Kuder-Richardson (Downie & Heath, 1979): $r_{ii} = (k / (k-1)) (1 - (\sum pq / s^2))$, em que k corresponde ao número de itens do teste (isto é, 125); p corresponde à proporção de respondentes que responderam corretamente aos itens 1 a 125; q corresponde a $1 - p$; e s^2 corresponde à variância do teste. Assim, na fórmula são somados os produtos pq para cada um dos 125 itens do teste, o que no presente caso resultou em 12,571. Considerando que a variância do teste para a presente amostra de 680 respondentes foi de 225,45 para pontuação simples, e de 414,57 para pontuação piso-teto, o coeficiente de consistência interna de Kuder-Richardson r_{ii} foi de 0,951 para pontuação simples, e de 0,977 para pontuação piso-teto. Tomando a pontuação padrão de acordo com as normas hispano-

americanas, a variância no teste foi de 197,34 e, logo, o coeficiente de consistência interna foi de 0,944.

De acordo com Downie & Heath (1979), os coeficientes de confiabilidade de testes estandardizados bem feitos tendem a ser altos, em cerca de 0,9 ou acima. Com base neste critério usual, portanto, é possível afirmar que a confiabilidade da presente tradução do *Teste de Vocabulário por Imagens Peabody* é, de fato, excelente.

Para expressar a confiabilidade de um teste, em vez de coeficientes de correlação, a estatística *erro padrão de mensuração* pode ser preferida. De acordo com Downie & Heath (1979), diferentemente de coeficientes de confiabilidade, tal estatística não é afetada pela amplitude dos escores da amostra, e é aproximadamente constante para amostras com diferentes variâncias. O erro padrão da mensuração corresponde ao desvio-padrão de uma amostra de pontuações de um respondente em relação à sua pontuação verdadeira. Para compreender bem tal conceito, é conveniente imaginar uma situação em que um mesmo teste fosse administrado repetidamente a um dado respondente. Em tal situação, as pontuações tenderiam a ser levemente diferentes umas das outras ao longo das testagens repetidas. Tomando a média de tais pontuações como uma estimativa da pontuação verdadeira, o erro padrão da mensuração poderia ser definido como o desvio-padrão dessas pontuações em relação à pontuação verdadeira. Quando a presença de efeitos de carreamento (i.e., *carry-over effects*) entre uma e outra testagem produz tendências não-aleatórias nas pontuações, é inadequado reaplicar o teste repetidamente, e faz-se necessário estimar o erro padrão de mensuração. Para tanto, a seguinte fórmula é recomendada (Downie & Heath, 1979):

$$s_e = s \sqrt{1 - r_{tt}}$$

em que s_e corresponde ao erro padrão de mensuração; s corresponde ao desvio-padrão do teste; e r_{tt} corresponde à confiabilidade do teste.

Como na presente testagem da amostra de 680 respondentes o desvio-padrão foi 15,01 para a pontuação simples, 20,36 para a pontuação piso-teto, e 14,05 para pontuação padrão conforme normas hispano-americanas; e, conforme relatado acima, a confiabilidade do teste foi de 0,952 quando avaliada em termos de pontuação simples, de 0,977 quando avaliada em termos de pontuação piso-teto, e de 0,944 quando avaliada conforme normas hispano-americanas; o erro-padrão de mensuração para pontuação simples corresponde a 3,29; para pontuação piso-teto, a 3,09, para normas hispano-americanas, a 3,32. Isto significa que a pontuação verdadeira de um dado respondente no teste não deve diferir mais do que 3,29 unidades de pontuação simples, ou do que 3,09 unidades de sua pontuação piso-teto, ou do que 3,32 unidades de sua pontuação conforme normas hispano-americanas. Mais especificamente, há uma probabilidade de 95 em 100 de que a pontuação verdadeira não difira da pontuação simples em mais que 6,58 unidades, que ela não difira da pontuação piso-teto em mais que 6,18 unidades, e que ela não difira da pontuação conforme normas hispano-americanas em mais de 6,65 unidades.

Quanto menor o erro padrão de mensuração de um teste, tanto mais confiável é este teste. Como a inspeção da fórmula demonstra, se um teste tivesse perfeita confiabilidade, isto é $r_{tt} = 1$, o erro-padrão de mensuração seria zero. Lord (1959) demonstrou que uma boa estimativa do erro padrão de mensuração pode ser obtida diretamente por meio da fórmula:

$$s_e = 0,432 \sqrt{k}$$

em que k corresponde ao número de itens do teste. Como PPVT tem 125 itens, o erro padrão de mensuração esperado a partir da fórmula de Lord corresponde a 4,83. Como o

erro-padrão de mensuração de PPVT calculado foi inferior ao erro-padrão esperado de um teste com tal número de itens, a presente tradução do teste PPVT pode ser considerada como sendo 47% mais confiável do que esperado quando avaliada via pontuação simples, 56% mais confiável do que esperado quando avaliada via pontuação piso-teto, e 23% mais confiável do que esperado quando avaliada via pontuação de acordo com normas hispano-americanas.

Discussão

O presente estudo forneceu uma tradução brasileira do *TVIP*, bem como normas fluminenses de desempenho de pré-escolares e escolares de primeira a oitava série nessa versão brasileira. Foram fornecidas normas em termos faixa etária de 6 a 18 anos, bem como em termos de nível de escolaridade de pré-escola até a oitava série. Tais normas foram obtidas a partir de aplicação coletiva do teste em sala de aula, em vez de a partir de aplicação individual. A opção pela aplicação coletiva foi feita a partir da opinião dos autores de que a aplicação coletiva é mais apropriada às condições de prática de ensino e de pesquisa de ensino no Brasil. Além disso, a adoção da aplicação coletiva possibilita simplificar a avaliação de desempenho dos examinandos, permitindo substituir os complexos critérios de piso e teto pelo simples cômputo da frequência de acertos. Assim, em vez de procurar obter a pontuação piso-teto, é necessário apenas obter a pontuação simples. O fato de que a aplicação coletiva oferece grande economia de recursos humanos e materiais, não apenas na fase de aplicação como também na de avaliação, pode contribuir para um emprego mais frequente do *TVIP* como instrumento de pesquisa de aprendizagem e desenvolvimento.

Além das vantagens de praticidade de aplicação e avaliação, a adaptação do teste para a sala de aula permite tirar vantagem de um contexto natural de avaliação nessa sala de

aula com o qual os examinandos já se encontram acostumados, que faz parte de seu dia-a-dia. Para os alunos pode ser mais natural assinalar em cadernos sob instrução de seus professores em sala de aula juntamente com seus colegas do que ser retirado da sala de aula e colocado sozinho numa sala isolada frente a um estranho numa situação artificial de avaliação. Como o teste se destina a uma população escolar e pré-escolar que está acostumada a avaliações coletivas de desempenho, a adaptação do teste para aplicação coletiva coaduna-se com as práticas no contexto escolar. No entanto, tal aplicação coletiva poderia ser pouco recomendável se houvesse discrepância entre a pontuação simples e a pontuação piso-teto. Os dados presentes indicam que há forte correlação positiva entre a pontuação simples e aquela proveniente da adoção daqueles critérios. Tal correlação torna equivalentes essas medidas para fins de avaliação. Isto permite prescindir da aplicação dos critérios de piso e teto que restringiam a aplicação do teste ao modo individual.

A evidência de equivalência entre as medidas de pontuação simples e de pontuação piso-teto parece diminuir nossa tendência a compartilhar do temor de Dunn, Padilla, Lugo & Dunn (1986) de que a aplicação de todos os itens do teste poderia não ser útil e mesmo invalidar o teste (p. 15). É possível que o temor de invalidação tenha fundamento em aplicação individual, já que a situação face-a-face entre examinando e examinador na aplicação individual pode contribuir para aumentar o desconforto frente a itens excessivamente fáceis ou difíceis ao examinando. Em aplicação individual, frente a itens excessivamente fáceis, o examinando poderia ter a sensação de ridículo, de que o examinador o estaria subestimando ou "fazendo pouco dele"; enquanto que, frente a itens excessivamente difíceis, o examinando poderia sentir-se envergonhado e constrangido por ter seus erros acompanhados tão de perto pelo examinador. Por outro lado, em aplicação coletiva há menos motivos para sentimentos de estar sendo subestimado, já que sempre há

alguém mais fraco para quem perguntas fáceis são apropriadas; nem envergonhado, já que há mais pessoas cometendo os mesmos erros frente aos mesmos itens.

Ainda assim, a noção de faixa discriminativa é útil e não deve ser desprezada. A faixa discriminativa pode ser definida como o segmento de itens que proporcionam o máximo de discriminação entre examinandos de um mesmo nível escolar ou de uma mesma faixa etária. Ela exclui os itens excessivamente fáceis e aqueles excessivamente difíceis. Não há porque supor que para aplicar a noção de faixa discriminativa seja necessário limitar o teste a aplicação individual. Em verdade, esta é mais uma das funções das tábuas de normatização. Assim, uma vantagem adicional da adaptação do teste a sala de aula, com aplicação coletiva e normatização por nível de escolaridade, é justamente que os respondentes pertencem precisamente ao mesmo nível de escolaridade. Isto permite que, mesmo em aplicação coletiva os examinandos sejam poupados dos extremos de facilidade e dificuldade. A partir da observação da Tábua 2 (para standardização em termos de nível de escolaridade a partir de pontuação simples), nota-se que a aplicação pode ser iniciada nos itens 8, 23, 40, 41, 55, 67, 84, 90, e 96 para a pré-escola e as séries primeira a oitava, respectivamente. Pode-se notar também que a aplicação deve ser interrompida nos itens 105, 108, 116, 120, 124 para as séries pré-escola, a quarta série, respectivamente. O mesmo pode ser feito em termos de idade. A partir da observação da Tábua 4 (para standardização em termos de faixa etária a partir de pontuação simples), nota-se que a aplicação pode ser iniciada no item 10 para crianças com seis e sete anos; e nos itens 24, 38, 44, 58, 65, 73, 88, 90, 93 para respondentes com oito a 16 anos de idade, respectivamente. Pode-se notar também que a aplicação deve ser interrompida nos itens 104, 110, 113, e 117 para as faixas etárias de 6, 7, 8, e 9 anos de idade, respectivamente. Tais diretrizes de aplicação são, a propósito, mais um dos usos da tabela de escores padronizados por nível de escolaridade.

O presente estudo não inclui uma comparação entre desempenhos em diferentes aplicações. Isto é, ele não comparou a aplicação individual com a aplicação coletiva. O estudo fez uma comparação entre métodos de avaliação diferentes de um único desempenho numa dada situação de aplicação coletiva. Isto é, o estudo comparou a pontuação simples com aquela baseada em critérios de piso e teto. O objetivo foi verificar se, na mesma situação de aplicação coletiva, a pontuação simples equivalia à pontuação piso-teto, ou seja, se ambas estariam correlacionadas entre si, e se seriam afetadas e não-afetadas igualmente pelas mesmas variáveis e nas mesmas medidas. Tendo demonstrado esses efeitos, o estudo buscou legitimar o uso da pontuação simples, facilitando assim o emprego do teste na prática e pesquisa educacional baseada em delineamentos de grupo

Ainda assim o estudo não comparou as duas pontuações em situações de aplicação diferentes, individual e coletiva. Não foi possível cruzar as duas formas de pontuação com as duas formas de aplicação. Isto não foi feito porque não seria possível aplicar o teste coletivamente respeitando os critérios de piso e teto. Em verdade, a adoção dos critérios de piso e base para a aplicação do teste inviabiliza a aplicação coletiva. Para comparar as duas formas de pontuação em aplicação individual seria preciso proceder ao reteste balanceado dos mesmos respondentes. A aplicação coletiva permite enorme economia de recursos. Ainda assim, a aplicação individual é a única situação que permite adotar os critérios de piso e teto já na aplicação do teste. De modo a saber se esta vantagem é ou não realmente importante seria necessário comparar o desempenho no teste inteiro com o desempenho apenas na faixa crítica do teste (isto é, de piso ao teto apenas). O presente estudo demonstrou apenas que a maneira de avaliar não faz muita diferença quando o teste inteiro é aplicado, mas ele não produziu qualquer dado comparando o efeito da aplicação do teste inteiro com a aplicação de faixa crítica apenas. Tanto quanto sabemos tais dados inexistem na literatura e aguardam pesquisa ulterior.

Referências bibliográficas

- Barlow, D.H. & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change*. (2nd ed). New York, N.Y.: Pergamon Press.
- Burgemeister, B., Blum, L. & Lorge, I. (1972). *Columbia Mental Maturity Scale* (3rd. ed.). New York, N.Y.: Harcourt, Brace, Jovanovich.
- Campbell, D.T. & Stanley, J.C. (1979). *Delineamentos experimentais e quase-experimentais de pesquisa*. São Paulo, S.P.: EPU-EDUSP.
- Campos, M.R. (1994). *Teste de Maturidade para a Leitura*. Rio de Janeiro, R.J.: Centro Editor de Psicologia Aplicada.
- Capovilla, F.C. (1993). Informática aplicada à neuropsicologia. Em N. Rodrigues e L. Mansur (Orgs). *Temas em Neuropsicologia*, vol 1, Sociedade Brasileira de Neuropsicologia, São Paulo, S.P., pp. 130-140.
- Capovilla, F.C. (1994 a). Informática em educação especial: comunicação na ausência de fala. *Proceedings of the XVII International School Psychology Colloquium e Anais do II Congresso Nacional de Psicologia Escolar, Campinas, S.P.*, pp. 155-158.
- Capovilla, F.C. (1994 b). Pesquisa e desenvolvimento de novos recursos tecnológicos e conceituais para educação especial. Em E.M.L.S. Alencar (Org.). *Tendências e Desafios da Educação Especial*. Série Atualidades Pedagógicas, MEC-SEESP, Brasília, D.F., pp. 196-212.
- Capovilla, F.C. (1996). Sistemas especialistas de multimídia em educação especial. Em L.R.O.P. Nunes (Org.). *Prevenção e intervenção educação especial*. Série Coletâneas da ANPEPP, vol 14. Rio de Janeiro, R.J.: Associação Nacional de Pesquisa e Pós-Graduação em Psicologia, pp. 124-150.
- Capovilla, F.C., Colomi, E.M., Nico, M.A. & Capovilla, A.G.S. (1995). Leitura em voz alta, tomada de ditado, manipulação fonêmica, e relações entre elas: Efeito de

características de palavras (frequência, regularidade, lexicalidade) e de nível de escolaridade. Em B. Damasceno e M.I.Coudry (Orgs.). *Temas em Neuropsicologia e Neurolinguística*. Série de Neuropsicologia, vol 4, Sociedade Brasileira de Neuropsicologia, Campinas, S.P., pp. 157-168.

- Capovilla, F.C., Gonçalves, M.J., Macedo, E.C., Duduchi, M. & Capovilla, A.G. (1996). Evidence of verbal processes in message encoding by cerebral-palsied using a picto-ideographic AAC system. *Proceedings of the 7th Biennial Conference of the International Association for Augmentative and Alternative Communication*. Vancouver, B.C., Canada, August, pp. 148-149.
- Capovilla, F.C., Guedes, M., Macedo, E.C., Thiers, V.O., Raphael, W.D., Capovilla, A.G.S. & Duduchi, M. (1995). Versões computadorizadas de testes de aptidão escolástica para deficientes motores: *Teste de Vocabulário por Imagens Peabody*, Teste de Figuras Invertidas, Escala de Maturidade Mental Columbia. *Resumos do II Congresso da Sociedade Brasileira de Neuropsicologia, Unicamp, Campinas, S.P.*, p. P17.
- Capovilla, F.C. & Macedo, E.C. (1994). Ferramentas de informática em pesquisa e prática psicopedagógica. *Proceedings of the XVII International School Psychology Colloquium e Anais do II Congresso Nacional de Psicologia Escolar, Campinas, S.P.*, pp. 121-125.
- Capovilla, F.C., Macedo, E.C., Capovilla, A.G.S., Duduchi, M. & Gonçalves, M.J. (no prelo). O uso de sistemas alternativos e facilitadores de comunicação para o tratamento e a melhoria da qualidade de vida de afásicos. *O Mundo da Saúde*, no prelo.
- Capovilla, F.C., Macedo, E.C., Capovilla, A.G.S., Raphael, W.D. & Gonçalves, M.J. (1996). O lugar da língua de sinais e dos sistemas de sinais na educação da criança surda. *O Mundo da Saúde*, 20(9), 292-297.
- Capovilla, F.C., Macedo, E.C. & Duduchi, M. (1994). Teste de atenção antecipatória de eventos como pré-requisito à comunicação de lesado cerebral incapaz de vocalizar e de

- qualquer controle voluntário sobre musculatura esquelética. *Resumos da XXIV Reunião Anual da Sociedade Brasileira de Psicologia*, Ribeirão Preto, S.P., p. 193.
- Capovilla, F.C., Macedo, E.C., Duduchi, M., Capovilla, A.G.S. & Gonçalves, M.J. (1996). "Brincar de ler": O computador no diagnóstico diferencial das dislexias. *O Mundo da Saúde*, 20(2), 87-89.
- Capovilla, F.C., Macedo, E.C., Duduchi, M., Capovilla, A.G., Raphael, W.D. & Guedes, M. (1996). UltraActive: Computerized multimedia expert AAC system. *Proceedings of the 7th Biennial Conference of the International Association for Augmentative and Alternative Communication*. Vancouver, B.C., August, pp. 467-468.
- Capovilla, F.C., Macedo, E.C., Duduchi, M., Gonçalves, M.J. & Capovilla, A.G. (1996). Home use of a computerized picto-syllabic-speaking AAC system in cerebral-palsy: preliminary data. *Proceedings of the 7th Biennial Conference of the International Association for Augmentative and Alternative Communication*. Vancouver, B.C., Canada, August, pp. 463-464.
- Capovilla, F.C., Macedo, E.C., Duduchi, M. & Guedes, M. (1996). Synthesized or digitized speech? That is the question. NoteVox: the answer. *Proceedings of the 7th Biennial Conference of the International Association for Augmentative and Alternative Communication*. Vancouver, B.C., Canada, August, pp. 465-466.
- Capovilla, F.C., Macedo, E.C., Duduchi, M., Raphael, W.D., Capovilla, A.G. & Guedes, M. (1996). Computerized tools for assessing scholastic progress in AAC users with severe motor impairments. *Proceedings of the 7th Biennial Conference of the International Association for Augmentative and Alternative Communication*. Vancouver, B.C., Canada, August, pp. 55-56.

- Capovilla, F.C., Macedo, E.C., Duduchi, M., Raphael, W.D., Guedes, M., Capovilla, A.G.S. & Gonçalves, M.J. (1996). Avaliação computadorizada de vocabulário e compreensão auditiva em crianças falantes ou não. *O Mundo da Saúde*, 20(1), 421-424.
- Capovilla, F.C., Macedo, E.C., Duduchi, M., Thiers, V.O., Capovilla, A.G.S. & Gonçalves, M.J. (1995). Como selecionar o melhor sistema de comunicação para seu paciente com déficit de fala? *O Mundo da Saúde*, 19(10), 350-352.
- Capovilla, F.C., Macedo, E.C. & Feitosa, M.D. (1994). Alternative input devices for computerized communication systems. *Abstracts of the VII Meeting of the International Society for Comparative Psychology*, São Paulo, S.P., p. 67.
- Capovilla, F.C., Macedo, E.C., Gonçalves, M.J., Capovilla, A.G.S., Raphael, W.D. & Duduchi, M. (1995). Computer systems for assessing reading, writing, and phonologic segmentation abilities. *Proceedings of the Third ECART: European Conference on the Advancement of Rehabilitation Technology*, Lisboa, Portugal, pp. 86-88.
- Capovilla, F.C., Macedo, E.C., Raphael, W.D., Capovilla, A.G.S., Gonçalves, M.J., Duduchi, M. & Guedes, M. (1995). Multimedia expert systems for cognitive evaluation of AAC system users in special education. *Proceedings of the Third ECART: European Conference on the Advancement of Rehabilitation Technology*, Lisboa, Portugal, pp. 89-91.
- Capovilla, F.C., Macedo, E.C., Raphael, W.D., Duduchi, M., Moreira, M.A.C., Gonçalves, M.J. & Capovilla, A.G.S. (1995). Multimedia expert systems for communication and education of the hearing impaired. *Proceedings of the Third ECART: European Conference on the Advancement of Rehabilitation Technology*, Lisboa, Portugal, pp. 83-85.
- Capovilla, F.C., Macedo, E.C., Seabra, A.G., Feitosa, M.D. & Thiers, V.O. (1994). Sistemas computadorizados para surdo-mudos baseados em língua de sinais:

- Comunicação via Logofone e ensino via Logofone Tutor. *Anais da II Jornada USP-SUCESU-SP de Informática e Telecomunicações*, São Paulo, S.P., pp. 363-372.
- Capovilla, F.C., Nunes, L., Capovilla, A.G., Raphael, W.D. & Gonçalves, M.J. (no prelo). Avaliação do desenvolvimento verbal da criança via *Teste de Vocabulário por Imagens Peabody*. *O Mundo da Saúde*, no prelo.
- Capovilla, F.C., Raphael, W.D., Capovilla, A.G., Guedes, M., Costa, C.E., Macedo, E.C., Duduchi, M., Aligieri, S., Santos, A., Viana, A.L. & Fuso, S.F. (1996). Sistema de multimídia para comunicação surdo-surdo e surdo-ouvinte em línguas brasileira e americana de sinais via redes de computador. *O Mundo da Saúde*, 20(3), 110-114.
- Dale, E. & Reichert, D. (1957). *Bibliography of vocabulary studies*. Columbus, OH: Ohio State University Bureau of Educational Research.
- DeRenzi, E., & Vignolo, I. (1962). The Token Test: A sensitive test to detect receptive disturbances in aphasics. *Brain*, 85, 665-687.
- DiSimoni, F. (1978). *The Token Test for children*. Boston, MA: Teaching Resources Corporation.
- Downie, N.M. & Heath, R.W. (1979). *Basic statistical methods*. 3rd. ed., New York, N.Y.: Harper & Row.
- Dunn, L.M. (1959). *Peabody Picture Vocabulary Test*. Circle Pines, MN: American Guidance Service.
- Dunn, L.M. & Dunn, L.M. (1981). *Peabody Picture Vocabulary Test - Revised*. Circle Pines, MN: American Guidance Service.
- Dunn, L.M., Padilla, E.R., Lugo, D.E. & Dunn, L.M. (1986). *Test de Vocabulário em Imágenes Peabody: Adaptación Hispanoamericana*. Circle Pines, MN: American Guidance Service.
- Edfeldt, A.W. (1955). Reading reversal and its relations to reading readiness. *Research Bulletin from the Institute of Education*, no. 1, University of Stockholm.
- Elliott, C.D. (1983). *The British ability scales. Manual 2: Technical and statistical information*. Windsor, U.K.: NFER-Nelson.
- Eysenck, M.W. & Keane, M.T. (1990). *Cognitive psychology: A student's handbook*. New York, N.Y.: Lawrence Erlbaum.
- Feitosa, M.D., Macedo, E.C., Capovilla, F.C., Seabra, A.G. & Thiers, V.O. (1994). Sistemas computadorizados de comunicação e de ensino para paralisia cerebral baseados na linguagem Bliss. *Anais da II Jornada USP-SUCESU-SP de Informática e Telecomunicações*, São Paulo, S.P., pp. 343-352.
- Gonçalves, M.J., Capovilla, F.C. & Macedo, E.C. (no prelo). A fonoaudiologia na era da informática e seu encontro com a comunicação alternativa e facilitadora. Em C.César e M.Lagrotta (Orgs.). *Tópicos em Fonoaudiologia*. São Paulo, S.P.: Editora Lovise.
- Gonçalves, M.J., Capovilla, F.C., Macedo, E.C., Feitosa, M. & Seabra, A.G. (1995). Quando falar não é possível: Uma alternativa. *Cadernos de Estudos Linguísticos*, 29, 49-56.
- Gonçalves, M.J., Macedo, E.C., Duduchi, M., Capovilla, A.G., Thiers, V.O. & Capovilla, F.C. (1995). Comunicação computadorizada a serviço da saúde e da qualidade de vida. *O Mundo da Saúde*, 19(4), 145-148.
- Gonçalves, M.J., Macedo, E.C., Duduchi, M. & Capovilla, F.C. (1995). Computerized Pictogram Ideogram Communication System for cerebral palsy: preliminary data. *Proceedings of the Third ECART: European Conference on the Advancement of Rehabilitation Technology*, Lisboa, Portugal, pp. 92-94.
- Guedes, M., Raphael, W.D., Capovilla, F.C., Macedo, E.C. & Duduchi, M. (1995). Avaliação computadorizada da iconicidade para brasileiros das figuras do *Teste de*

- Vocabulário por Imagens Peabody. Resumos da XXV Reunião Anual da Sociedade Brasileira de Psicologia, Ribeirão Preto, S.P., p. 265.*
- Kirk, R.E. (1982). *Experimental design: Procedures for the behavioral sciences*. 2nd ed. Belmont, CA: Brooks-Cole.
- Kunz, E.R. (1979). *Teste de prontidão para a leitura*. Rio de Janeiro, R.J.: Centro Editor de Psicologia Aplicada.
- Macedo, E.C., Capovilla, F.C., Gonçalves, M.J., Seabra, A.G., Thiers, V.O. & Feitosa, M.D. (1994). Adaptando um sistema computadorizado pictográfico para comunicação em paralisia cerebral tetra-espástica. *Anais da II Jornada USP-SUCESU-SP de Informática e Telecomunicações*, São Paulo, S.P., pp. 353-362.
- McKeon, M.G., Beck, I.L., Omanson, R.C. & Pople, M.T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly*, 20, 522-535.
- Nunes, L.R.O.P., Capovilla, F.C., Martins, I., Nunes, D., Nogueira, D., Bernart, A.B., Ribeiro, A. & Carvalho, V. (1995). Normatização brasileira do *Teste de Vocabulário por Imagens Peabody*: dados de uma amostra fluminense. *Resumos do V Congresso Interno do Núcleo de Pesquisa em Neurociências e Comportamento*. Universidade de São Paulo, São Paulo, S.P., p. 5.
- Raphael, W.D., Guedes, M., Capovilla, F.C., Macedo, E.C., Duduchi, M. & Capovilla, A.G.S. (1995). Sistema de multimídia para avaliação computadorizada de vocabulário de pessoas não-falantes: *Teste de Vocabulário por Imagens Peabody - Versão computadorizada*. *Resumos do V Congresso Interno do Núcleo de Pesquisa em Neurociências e Comportamento*, Universidade de São Paulo, São Paulo, S.P., p. 16.
- Robertson, J.R. & Eisenberg, J.L. (1981). *Technical supplement to the Peabody Picture Vocabulary Test - Revised*. Circle Pines, MN: American Guidance Service.
- Sternberg, R.J. (1985). *Beyond I.Q.: A triarchic theory of human intelligence*. Cambridge, MA: Cambridge University Press.
- Thiers, V.O. & Capovilla, F.C. (1996). Alternative communication in cerebral-palsy: Evaluation of variables that control the search for *Blissymbols* in communication boards. *Proceedings of the 7th Biennial Conference of the International Association for Augmentative and Alternative Communication*. Vancouver, B.C., Canada, August, pp. 60-61.
- Thiers, V.O., Capovilla, F.C., Macedo, E.C., Feitosa, M.D. & Seabra, A.G. (1994). Aplicação do *software Sonda* para análise diferencial de iconicidade em sistemas de comunicação para pacientes neurológicos. *Anais da II Jornada USP-SUCESU-SP de Informática e Telecomunicações*, São Paulo, S.P., pp. 373-382.
- Warren S.F. & Kaiser, A.P. (1986). Generalization of treatment effects by young language-delayed children: A longitudinal analysis. *Journal of Speech and Hearing Disorders*, 51, 239-251.
- Wechsler, D. (1991). *David Wechsler test de inteligencia para niños*. Buenos Aires, Argentina: Editorial Paidós.