

Universidade de São Paulo
Instituto de psicologia

Andréa Jacote

**Banco para avaliar linguagem, controlando: univocidade de figuras,
familiaridade e decifrabilidade de escrita; cifrabilidade de fala ouvida; e
legibilidade, audibilizabilidade e cifrabilidade de fala vista**

Volume 1

**São Paulo
2015**

Andrea Jacote

Banco para avaliar linguagem, controlando: univocidade de figuras, familiaridade e decifrabildade de escrita; cifrabildade de fala ouvida; e legibilidade, audibilizabilidade e cifrabildade de fala vista

Dissertação apresentada ao Instituto de Psicologia da Universidade de São Paulo como parte dos requisitos para a obtenção do grau de Mestre em Psicologia

Área de Concentração:
Psicologia Experimental

Orientador: Prof. Titular
Dr. Fernando César Capovilla

Volume 1

São Paulo
Janeiro de 2015

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

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

Lima, Andréa Jacote.

Banco para avaliar linguagem, controlando: univocidade de figuras, familiaridade e decifrabildade de escrita; cifrabilidade de fala ouvida; e legibilidade, audibilizabilidade e cifrabilidade de fala vista / Andréa Jacote Lima; orientador Prof. Dr. Fernando César Capovilla. -- São Paulo, 2014. 881f.

Dissertação (Mestrado – Programa de Pós-Graduação em Psicologia. Área de Concentração: Psicologia Experimental) – Instituto de Psicologia da Universidade de São Paulo.

- Univocidade 2. Familiaridade 3. Ortografia
- 4. Cifrabilidade 5. Decifrabildade 6. Fanerolaliema 7. Grafema 8.
- Fonema I. Título.

BF455

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE PSICOLOGIA

*Dissertação apresentada ao Instituto de Psicologia da
Universidade de São Paulo como parte dos requisitos para
a obtenção do grau de MESTRE EM PSICOLOGIA*

Andréa Jacote

Área de Concentração: Psicologia Experimental

Aprovada em: ____/____/____

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Agradecimentos

Agradeço em primeiro lugar a Deus por me proporcionar a oportunidade de realizar um dos maiores sonhos da minha vida, mesmo com todas as improbabilidades de consegui-lo.

A minha Mãe (in memoriam) que foi a responsável pela minha auto confiança e pela minha determinação, através do amor que sempre demonstrou e pela pessoa maravilhosa que sempre foi.

A minha família querida e amada, minhas irmãs Adriana e Aline que aguentaram todo o desgaste proporcionado pela minha falta de tempo e excesso de trabalho com paciência e carinho me ajudaram a superar minhas crises e fraquezas não deixando que eu desistisse desse sonho que, em certos momentos, se tornou quase irrealizável devido à sua extensão e complexidade, e ao grande amor da minha vida, a minha filha Melissa, que, apesar de ter a mãe ausente em momentos muito importantes de sua vida, nunca cobrou e sempre esteve ao meu lado incentivando e acreditando no meu amor.

Ao meu orientador Prof. Dr. Fernando César Capovilla, que acreditou na minha capacidade de superar dificuldades, mesmo quando eu mesma já não acreditava, e com toda sua sabedoria me guiou pelo mundo do conhecimento me transformando em uma pessoa bem melhor do que era antes de conhecê-lo.

Sem querer cometer nenhum tipo de injustiça deixando de mencionar alguém, quero agradecer a todos os meus amigos que acompanharam essa minha jornada e em momentos cruciais através de gestos e palavras me fizeram acreditar que seria possível, contribuindo assim para que eu conseguisse chegar ao final dessa grande etapa da minha vida.

Índice

1	Resumo	7
2	Palavras-chave	7
3	Abstract	8
4	Keywords	8
5	Extend Abstract	9
6	Breve perspectiva geral	10
7	Introdução	10
8	Univocidade média da figura	12
9	Decifrabildade <i>GrafEma-FonEma</i> média da palavra escrita	13
10	Palavras faladas recebidas por audição	15
11	Cifrabildade média <i>FonEma-GrafEma</i> da palavra ouvida	15
12	Palavras faladas recebidas por leitura orofacial visual - Legibilidade orofacial média da palavra falada recebida por visão	18
13	Modelos de legibilidade orofacial	19
14	Implementação dos Modelos de Legibilidade Orofacial: Implementação Simples (IS) <i>versus</i> Implementação com Valor Acentual (IVA)	24
15	Apresentação e descrição pormenorizada dos quadros fonéticos dos segmentos que ocorrem no Português brasileiro, conforme Capovilla e Sousa-Sousa (em preparação a, b, c)	26
16	Cifrabildade média <i>FaneroLaliEma-GrafEma</i> da palavra lida orofacialmente por visão	37
17	Audibilizabilidade média <i>FaneroLaliEma-FonEma</i> da palavra lida orofacialmente por visão	76
18	Familiaridade ortográfica das palavras	99
19	Breve apresentação do banco de 971 entradas	103
20	Descrição do banco de figuras	109
21	Referências bibliográficas	111
22	Banco de figuras e palavras	116
23	Anexo 1: Tabelas de Cifrabildade e de Decifrabildade.	364
26	Anexo 2: Tabelas Decifrabildade <i>GrafEma-FonEma</i> - (IS)	396
27	Anexo 3: Tabelas Cifrabildade <i>FonEma-GrafEma</i> - (IS)	464
28	Anexo 4: Tabelas Cifrabildade <i>FonEma-GrafEma</i> - (IVA)	531
29	Anexo 5: Tabelas Cifrabildade <i>FaneroLaliEma-GrafEma</i> - (IS)	600
30	Anexo 6: Tabelas Cifrabildade <i>FaneroLaliEma-GrafEma</i> - (IVA)	668
31	Anexo 7: Tabelas Audibilizabilidade <i>FaneroLaliEma-Fonema</i> - (IS)	736
32	Anexo 8: Tabelas Audibilizabilidade <i>FaneroLaliEma-Fonema</i> - (IVA)	804
33	Anexo 9: Sub-banco: Entradas lexicais excluídas do Banco: Itens 1 a 29 (figuras e correspondentes palavras ou letras) que foram removidos do banco	872

JACOTE, A. L. (2015). **Banco para avaliar linguagem, controlando univocidade de figuras, familiaridade e decifrabilidade de escrita; cifrabilidade de fala ouvida; e legibilidade, audibilizabilidade e cifrabilidade de fala vista**

1. Resumo

Esta dissertação de mestrado apresenta um banco de figuras e palavras. O banco objetiva servir para a aumentar a validade e precisão dos instrumentos de avaliação, bem como a eficácia dos materiais instrucionais para desenvolvimento de linguagem. Este banco contém 971 entradas lexicais. Cada entrada contém uma figura e seu correspondente nome escrito. A figura é analisada em termos de grau de univocidade (grau de concordância na nomeação). O nome da figura é analisado separadamente em três formas: palavra escrita visível, palavra falada audível, e palavra falada visível (lida orofacialmente). Palavras escritas visíveis são compostas de grafemas. São analisadas em termos de seu grau de familiaridade e reconhecibilidade (grau em que pode ser lida via rota lexical) e decodificabilidade (grau em que pode ser lida pela rota perilexical ou fonológica). Palavras ouvidas audíveis são compostas de fonemas. São analisadas em termos do grau de cifrabilidade (grau de facilidade com que podem ser escritas via rota perilexical). Palavras faladas vistas são compostas de fanerolaliemas. São analisadas em termos do grau de legibilidade orofacial (grau em que podem ser compreendidas apenas pela visão), audibilizabilidade (grau em que a imagem auditiva dos fonemas pode ser evocada por fanerolaliemas durante a leitura orofacial visual), e cifrabilidade (grau de facilidade com que podem ser escritas via rota perilexical). O banco é composto de 971 entradas lexicais, cada qual composta de uma figura (à esquerda) e de vários dados pertinentes ao seu nome correspondente (à direita). O lado direito da entrada é composto de seis campos. O Campo 1 fornece o nome da figura escrito em dois alfabetos: alfabeto romano e alfabeto fonético internacional. Ele também fornece a categoria semântica à qual pertence a palavra. O Campo 2 fornece o número da figura (para indexar todas as 971 figuras do banco). O Campo 3 fornece a univocidade da figura numa escala de 0-100 pontos separadamente para crianças de 2, 3, 4, 5, 6 7-10 anos de idade, bem como para adultos. O Campo 4 fornece a palavra escrita visível tal como analisada em termos de seu grau de familiaridade ou reconhecibilidade (grau em que pode ser lida via rota lexical) numa escala de 1-9 pontos, separadamente para crianças de 5º ano, 4º ano, 3º ano, 2º ano, e 1º ano. Nessa escala, 5 corresponde à média, 6 a 1 erro-padrão (EP) acima da média, 7 a 2 EP acima da média e assim por diante até 9; ao passo que 4 corresponde 1 EP abaixo da média, 3 a 2 EP abaixo da média, e assim por diante até 1. O Campo 5 é composto de quatro linhas, cada qual dividida em quatro colunas. A Linha 1 fornece o grau de decifrabilidade (grau com que pode ser lida pela rota perilexical) da palavra escrita visível, numa escala de 0-1. A Linha 2 fornece o grau de cifrabilidade da palavra ouvida (grau com que pode ser escrita pela rota perilexical), numa escala de 0-1. A Linha 3 fornece o grau de audibilizabilidade da palavra falada lida orofacialmente (grau com que sequência de fanerolaliemas pode ser convertida em sequência de fonemas), numa escala de 0-1. A Linha 3 fornece o grau de cifrabilidade da palavra falada lida orofacialmente (grau com que sequência de fanerolaliemas pode ser convertida em sequência de grafemas), numa escala de 0-1. Cada palavra é dividida em suas colunas. Cada coluna fornece os dados referentes à linha em questão em uma de quatro formas diferentes. Nas Colunas 1 e 2, dados consistem na média das razões independente da incidência. Nas Colunas 3 e 4, dados consistem na média das razões ponderada por incidência diferencial. Nas Colunas 1 e 3 os dados consistem na média das razões independentemente da tonicidade da fala (seja ouvida ou vista) na pronúncia. Nas Colunas 2 e 4, os dados consistem na média das razões ponderada pela tonicidade diferencial da fala (seja ouvida ou vista) na pronúncia. Por exemplo, a Linha 1 fornece o grau de decifrabilidade grafema-fonema da palavra escrita visível. Na Coluna 1 decodificabilidade é calculada como mera média de razões independente da incidência ou tonicidade. Na Coluna 2 decodificabilidade é calculada como média das razões independente da incidência mas ponderada pela tonicidade. Na Coluna 3 decodificabilidade é calculada como média de razões ponderadas em termos de incidência mas independente de tonicidade. Na Coluna 4 decodificabilidade é calculada como média de razões ponderadas em termos de incidência e de tonicidade. O Campo 6 fornece o grau de legibilidade orofacial da fala vista, numa escala de 0-1. O grau de legibilidade orofacial é apresentado em quatro formas. Nas Colunas 1 e 2 ela se encontra calculada segundo o modelo Dória; nas Colunas 3 e 4 ela se encontra calculada segundo o modelo Fonético-Articulatório. Nas Colunas 1 e 3 ela é calculada independentemente da tonicidade da pronúncia; nas Colunas 2 e 4 ela é calculada de modo ponderado pela tonicidade diferencial da pronúncia.

2. Palavras-chave: Univocidade, familiaridade, ortografia, cifrabilidade, decifrabilidade, fanerolaliema, grafema, fonema.

JACOTE, A. L. (2015). **Language assessment sourcebook with control upon degree of picture univocity, print recognizability and decodibility, audible speech encodibility, and visible speech legibility, audibilizability and encodibility**

3. Abstract

This master's thesis presents a new sourcebook aimed at increasing the validity and precision of language assessment tools, as well as the efficacy of instructional materials for language development. The sourcebook contains 971 lexical entries. Each entry contains a picture and its corresponding written name. The picture is analyzed in terms of its degree of univocity (i.e., picture naming agreement). The picture name is analyzed separately in three forms: visual written word, auditory spoken word, and visual spoken word (i.e., speechreading). Visual written word is made of graphemes. It is analyzed in terms of its degree of both: familiarity or recognizability (i.e., the degree to which it is suitable to be read via lexical reading route) and decodibility (i.e., the degree to which it is suitable to be read via perilexical reading route). Auditory spoken word is made of phonemes. It is analyzed in terms of its degree of encodibility (i.e., the degree to which it may be suitable for writing or spelling via perilexical spelling route). Visual spoken word is made of visemes. It is analyzed in terms of its degree of: speechreadability (i.e., the degree to which it may be understood via visual speechreading), audibilizability (i.e., the degree to which the auditory imagery of phonemes can be evoked by mouthshapes or visemes during speechreading), and encodibility (i.e., the degree to which it is suitable to be written or spelled correctly via perilexical route). The sourcebook is made of 971 lexical entries. Each entry is made of a picture (on the left) and several data pertaining to its corresponding name (on the right). The right side of the entry is made of six areas. The first area provides the picture name as it is written in both alphabets: the Roman alphabet (orthographic form) and the International Phonetic Alphabet. It also provides the semantic category to which the word belongs. The second area provides the picture number (for indexing all the 971 pictures of the sourcebook). The third area provides the picture univocity in a 0-100 scale for children aged: 2 years, 3 years, 4 years, 5 years, 6 years, 7 to 10 years, as well as for adults. The fourth area provides the visual written word as it is analyzed in terms of its degree of familiarity or recognizability (i.e., the degree to which the written word is suitable to be read via lexical reading route) in a 1-9 point scale, for children from 5th grade, 4th grade, 3rd grade, 2nd grade, and 1st grade. In such a scale, 5 corresponds to the mean, 6 is the mean plus 1 standard error, 7 is the mean plus 2 standard errors and so forth until 9, whereas 4 corresponds to the mean minus 1 standard error, 3 corresponds to the mean minus 2 standard errors and so forth until 1, which corresponds to the mean minus 4 standard errors. The fifth area is made of four lines. Each line is divided into four columns. The first line provides the visual written word degree of decodibility (i.e., the degree to which it is suitable to be read via perilexical reading route) in a 0-1 scale. The second line provides the auditory spoken word degree of encodibility (i.e., the degree to which it may be suitable for writing or spelling via perilexical spelling route) in a 0-1 scale. The third line provides the visual spoken word degree of audibilizability (i.e., the degree to which the auditory imagery of phonemes can be evoked by mouthshapes or visemes during speechreading) in a 0-1 scale. The fourth line provides the visual spoken word degree of encodibility (i.e., the degree to which it is suitable to be written or spelled correctly via perilexical route) in a 0-1 scale. Each line is divided into four columns. Each column presents the data pertaining to the line in question in 1 of 4 different forms. In the first and second columns the data consist of the mean of the ratios regardless of incidence. In the third and fourth columns the data consist of the mean of the ratios weighted by differential incidence. In the first and third columns the data consist of the mean of the ratios regardless of tonicity of speech (either auditory or visual) in pronunciation. In the second and fourth columns the data consist of the mean of the ratios weighted by differential tonicity of speech (either auditory or visual) in pronunciation. For instance the first line provides the visual written word degree of decodibility (i.e., grapheme to phoneme decoding). In the first column decodibility is calculated as a mere mean of the ratios regardless of either incidence or tonicity. In the second column decodibility is calculated as a mean of the ratios regardless of incidence but weighted in terms of tonicity. In the third column decodibility is calculated as a mean of the ratios weighted in terms of incidence but regardless of tonicity. In the fourth column it is calculated as a mean of the ratios weighted in terms of both incidence and tonicity. The sixth area provides the visual spoken word degree of speechreadability (i.e., the degree to which it may be understood via visual speechreading) in a 0-1 scale. The speechreadability is presented in 1 of 4 different forms. In the first and second columns, the speechreadability is calculated according to Doria's model. In the third and fourth columns it is calculated according to a phonetic model. In the first column and third columns it is calculated regardless of tonicity in pronunciation. In the second and fourth columns it is calculated in a way that is weighted by the differential tonicity in pronunciation.

4. Keywords: Univocity, familiarity, orthography, encoding, decoding, viseme, grapheme, phoneme

5. Extended Abstract

The present master's thesis presents a new sourcebook aimed at increasing the validity and precision of language assessment tools, as well as the efficacy of instructional materials for language development. The sourcebook contains 971 lexical entries. Each entry contains a picture and its corresponding written name. The picture is analyzed in terms of its degree of univocity (i.e., picture naming agreement). The picture name is analyzed separately in three forms: visual written word, auditory spoken word, and visual spoken word (i.e., speechreading). Visual written word is made of graphemes. It is analyzed in terms of its degree of both: familiarity or recognizability (i.e., the degree to which it is suitable to be read via lexical reading route) and decodibility (i.e., the degree to which it is suitable to be read via perilexical reading route). Auditory spoken word is made of phonemes. It is analyzed in terms of its degree of encodibility (i.e., the degree to which it may be suitable for writing or spelling via perilexical spelling route). Visual spoken word is made of visemes. It is analyzed in terms of its degree of: speechreadability (i.e., the degree to which it may be understood via visual speechreading), audibilizability (i.e., the degree to which the auditory imagery of phonemes can be evoked by mouthshapes or visemes during speechreading), and encodibility (i.e., the degree to which it is suitable to be written or spelled correctly via perilexical route).

The sourcebook is made of 971 lexical entries. Each entry is made of a picture (on the left) and several data pertaining to its corresponding name (on the right). The right side of the entry is made of six areas. The first area provides the picture name as it is written in both alphabets: the Roman alphabet (orthographic form) and the International Phonetic Alphabet. It also provides the semantic category to which the word belongs. The second area provides the picture number (for indexing all the 971 pictures of the sourcebook). The third area provides the picture univocity in a 0-100 scale for children aged: 2 years, 3 years, 4 years, 5 years, 6 years, 7 to 10 years, as well as for adults. The fourth area provides the degree of familiarity or recognizability (i.e., the degree to which the written word is suitable to be read via lexical reading route) of the visual written word, in a 1-9 point scale, for children from 5th grade, 4th grade, 3rd grade, 2nd grade, and 1st grade. In such a scale, 5 corresponds to the mean, 6 is the mean plus 1 standard error, 7 is the mean plus 2 standard errors and so forth until 9, whereas 4 corresponds to the mean minus 1 standard error, 3 corresponds to the mean minus 2 standard errors and so forth until 1, which corresponds to the mean minus 4 standard errors.

The fifth area is made of four lines. Each line is divided into four columns. The first line (decif graf-fon) provides the degree of decodibility (i.e., the degree to which it is suitable to be read via perilexical reading route) of the visual written word in a 0-1 scale. The second line (cifrab fon-graf) provides the degree of encodibility (i.e., the degree to which it may be suitable for writing or spelling via perilexical spelling route) of the auditory spoken word, in a 0-1 scale. The third line (corr fanlal-fon) provides the degree of audibilizability (i.e., the degree to which the auditory imagery of phonemes can be evoked by mouthshapes or visemes during speechreading) of the visual spoken word in terms of its in a 0-1 scale. The fourth line (cifrab fanlal-graf) provides the visual spoken word as it is analyzed in terms of its degree of encodibility (i.e., the degree to which it is suitable to be written or spelled correctly via perilexical route) in a 0-1 scale.

Each line is divided into four columns. Each column presents the data pertaining to the line in question in 1 of 4 different forms (Br-IS, Br-IVA, Pond-IS, Pond-IVA). In the first column (Br-IS) and in the second column (Br-IVA) the data consist of the mean of the ratios regardless of incidence (Br). In the third column (Pond-IS) and fourth column (Pond-IVA) the data consist of the mean of the ratios weighted by differential incidence (Pond). In the first column (Br-IS) and third column (Pond-IS) the data consist of the mean of the ratios regardless of tonicity of speech (either auditory or visual) in pronunciation (IS). In the second column (Br-IVA) and fourth column (Pond-IVA) the data consist of the mean of the ratios weighted by differential tonicity of speech (either auditory or visual) in pronunciation (IVA).

For instance the first line provides the visual written word as it is analyzed in terms of its degree of decodibility (i.e., grapheme to phoneme decoding). In the first column (Br-IS) decodibility is calculated as a mere mean of the ratios regardless of either incidence (Br) or tonicity (IS). In the second column (Br-IVA) decodibility is calculated as a mean of the ratios regardless of incidence (Br) but weighted in terms of tonicity (IVA). In the third column (Pond-IS) decodibility is calculated as a mean of the ratios weighted in terms of incidence (Pond) but regardless of tonicity (IS). In the fourth column (Pond-IVA) decodibility is calculated as a mean of the ratios weighted in terms of both incidence (Pond) and tonicity (IVA).

The sixth area provides the degree of speechreadability (i.e., the degree to which it may be understood via visual speechreading) of the visual spoken word, in a 0-1 scale. The speechreadability is presented in 1 of 4 different forms (MD-IS, MD-IVA, MFA-IS, MFA-IVA). In the first column (MD-IS) and second column (MD-IVA), the speechreadability is calculated according to Doria's model (MD). In the third column (MFA-IS) and fourth column (MFA-IVA) the speechreadability is calculated according to a phonetic model (MFA). In the first column (MD-IS) and third column (MFS-IS) the speechreadability is calculated regardless of tonicity in pronunciation (IS). In the second column (MD-IVA) and fourth column (MFA-IVA) the speechreadability is calculated in a way that is weighted by the differential tonicity in pronunciation (IVA).