

Instruments for assessing receptive and expressive vocabularies in Brazilian Portuguese and Libras

Fernando Cesar Capovilla¹ & Miriam Damazio¹

¹ Institute of Psychology, University of Sao Paulo, São Paulo, Brasil

Correspondence: Fernando Cesar Capovilla
E-mail address: fcapovilla5@gmail.com

Abstract

The paper describes standardized instruments that researchers and practitioners can use for assessing receptive vocabulary development and expressive vocabulary development in both Brazilian Portuguese and Brazilian Sign Language (Libras) in both hearing and deaf children from 2 to 14 years of age. It establishes comparisons among such instruments based on their characteristics and indications, and summarizes some data on their validity, reliability and norms. Norms provided herein permit distinguishing among five talker groups (very late talkers, late talkers, on-time talkers, early talkers, very early talkers) at each of five age levels (i.e., from 1 to 5 years of age), as well as five listener groups (very late listeners, late listeners, on-time listeners, early listeners, very early listeners) at each of nine school grade levels (i.e., from Kindergarten to 8th grade). Norms provided herein also permit distinguishing among five sign receiver deaf groups (very late receivers, late receivers, on-time receivers, early receivers, very early receivers) at each of seven school grade levels of deaf education (i.e., from 2nd to 8th grade).

Keywords: vocabulary, speech comprehension, speech production, oral language, lexical development, PPVT

REVIEW

Introduction

Standardized instruments for assessing receptive and expressive vocabulary development in Brazilian Portuguese and Brazilian Sign Language

There are two instruments for assessing expressive vocabulary (i.e., word production) development in Brazil:

(1) The Brazilian version of the *Language Development Survey* (LDS: Rescorla, 1989), which is called *Lista de Avaliação de Vocabulário Expressivo* (LAVE: Capovilla & Capovilla, 1997);

(2) The original *University of Sao Paulo Expressive Vocabulary Test* (TVExp-Usp: Capovilla & Damazio, 2011).

There are four tests for assessing receptive vocabulary development in Brazil, three of them for assessing word comprehension of Brazilian Portuguese, and one for assessing sign comprehension of Brazilian Sign Language (Libras) signs.

For assessing word comprehension in Brazilian Portuguese, one can use three instruments:

(1) The Brazilian version of the *Peabody Picture Vocabulary Test* (PPVT: Dunn & Dunn, 1997), which is called *Teste de Vocabulário por Imagens Peabody* (TVIP: Capovilla & Capovilla, 1997);

(2) The *University of Sao Paulo Picture Vocabulary Test*, which is called *Teste de Vocabulário por Figuras USP* (TVF-Usp: Capovilla & Prudencio, 2006);

(3) The *Auditory Vocabulary Test*, which is called *Teste de Vocabulário Auditivo* (TVAud: Capovilla, Negrão, & Damazio, 2011).

For assessing sign comprehension in Brazilian Sign Language (Libras), one can use the *Libras Sign Comprehension Test* (which is called *Teste de Vocabulário Receptivo de Sinais de Libras* (TVRSL: Capovilla, Capovilla, Viggiano, & Bidá, 2004).

Instruments for assessing speech production in Brazil

The Brazilian version of the *Language Development Survey: Lista de Avaliação de Vocabulário Expressivo* (LAVE)

Language Development Survey (LDS: Rescorla, 1989) constitutes a word checklist. Its Brazilian version is called *Lista de Avaliação de Vocabulário Expressivo* (LAVE). It has been validated and standardized by Capovilla and Capovilla (1997) for middle-class children from 2 to 6 years of age. The inventory LAVE consists of two parts. The first part is a questionnaire that gathers data on children and their families. The second

part consists of a checklist of 309 words belonging to 14 semantic classes. Such a word checklist was conceived to be filled up by mothers who are to mark the words that the child uses to speak spontaneously. Capovilla and Capovilla (1997) evaluated 238 children from 2 to 6 years of age (103 from public schools, and 135 from private schools).

ANOVA showed that LAVE scores were a positive monotonic function of examinees' school grade level, $F(3, 109) = 7.03, p = .000$ and age in both years, $F(4, 120) = 3.44, p = .011$ and semesters, $F(7, 117) = 2.9, p = .008$. Bonferroni pair wise comparisons revealed that expressive vocabulary increased significantly from age 2 (230 words) to age 3 (270 words), from then to age 4 (290 words), and from then to age 5 (300 words). Based on such data, norms (i.e., standardization tables) were generated. In addition, results showed that school type (public *versus* private) also had effect: Private school examinees scored higher than did public school ones, $F(1, 161) = 4.87, p = .029$. Regression analysis of LAVE scores on word comprehension scores revealed positive correlation, $r = .36, F(1, 120) = 17.5, p = .000$. That is: children who scored higher in word production also scored higher in word comprehension.

Pedromônico, Affonso, and Sãnudo (2002) used LAVE to evaluate word

production in 30 lower-middle-class children (17 boys and 13 girls) from 22 to 36 months of age. Results showed an increase in word production as a function of age for both boys and girls. Children spoke on average 195 words. Girls spoke 43 words more than the boys did. Girls also spoke two more words per phrase than the boys did. Prevalent semantic classes were people, body parts, actions, home, and adjectives. Results showed the importance of LAVE as a means for using information provided by caretakers in early education for characterizing expressive vocabulary in Brazilian toddlers, and for helping to identify children in need for intervention designed to increase expressive language skills.

Ferracini, Capovilla, Dias, and Capovilla (2006) used LAVE to assess expressive language skills in 122 children from 3 to 5 years of age. ANOVA revealed that LAVE scores increased as a function of age. They also used TVIP to assess receptive language skills in the same children. Intelligence scores as assessed by *Columbia Mental Maturity Scale* (Burgemeister, Blum, & Lorge, 1971) were used as covariant. ANCOVA of language skills (LAVE scores and TVIP scores) as function of age (years of age), having intelligence level as covariant, revealed that word comprehension scores and word production scores increased

significantly from 3 years to 4 years of age, and from then to 5 years of age. LAVE average score was 253 points. This study suggests that LAVE is a fairly precise instrument: Cronbach alpha coefficient was .98; and split-half Spearman-Brown coefficient was .81.

The University of Sao Paulo Expressive Vocabulary Test: *Teste de Vocabulário Expressivo Usp (TVExp-Usp)*

The original version of the *University of Sao Paulo Expressive Vocabulary Test (TVExp-Usp)* was published originally by Capovilla and Damazio (2011). It evaluates word production skills in children from 18 months of age to 5 years and 11 months of age. TVExp-Usp contains 100 items, each one consisting of a line drawing to be named orally by the examinee. The test is implemented in a 100 page picture book in A5 size paper having one picture on each page. The examiner sits side by side of the child and holds the picture book. The examiner points at one picture, and asks the child to speak the name of it. Figure 1, top row, illustrates four of the 100 pictures of TVExp-Usp to be named by children, Test administration typically lasts 7 minutes for 4-6 year-old children, 15 minutes for 2-3 year-old children, and 30 minutes for 18 month-old children.

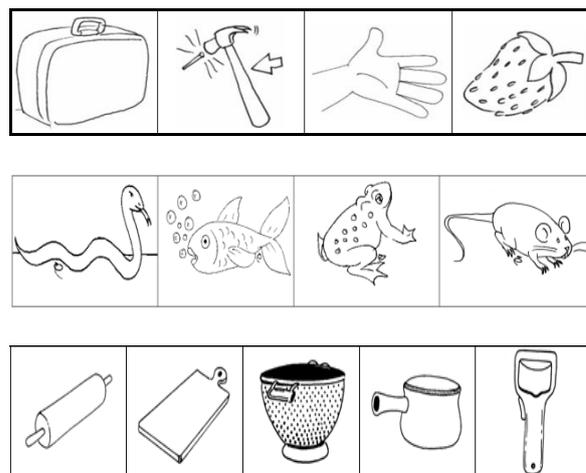


Figure 1. Top row: Four of the 100 pictures of the *Expressive Vocabulary Test (TVExp-Usp)*. Middle row: One of the 139 test items of the *Picture Vocabulary Test (TVF-Usp)*. Bottom row: One of the test items of the *Auditory Vocabulary Test (TVAud)*.

Capovilla and Damazio (2011) evaluated 332 children from nursery school to kindergarten. The sample comprised 61 1-year old children, 107 2-year old ones, 81 3-year old ones, 41 4-year old ones, and 42 5-year old ones. ANOVA showed that Expressive Vocabulary scores increased significantly with age, $F(4, 327) = 822.35, p < .000, r^2 = .91$. Bonferroni pair wise comparison revealed Expressive Vocabulary scores increased from ages 1 to 2, from 2 to 3, and from 3 to 4. Fisher LSD pair wise comparison also revealed that Expressive

REVIEW

Years of age	Very late talkers -3SD to -2SD	Late talkers -2SD to -1SD	On-time talkers -1SD to +1SD	Early talkers +1SD to +2SD	Very early talkers +2SD to +3SD
1 st	22	23	24 - 27	28	29
2 nd	25 - 36	37 - 49	50 - 76	77 - 89	90 - 100
3 rd	74 - 76	77 - 80	81 - 88	89 - 92	93 - 96
4 th	81 - 84	85 - 87	88 - 96	97 - 100	---
5 th	91 - 92	93 - 94	95 - 98	99 - 100	---

Table 1. *TVExp-Usp norms: Number of pictures (from 1 to 100) that are named by children from each of the five skill levels (very late talking, late talking, on-time talking, early talking, very early talking) for each of the five age levels (1 to 5 years of age).*

Vocabulary scores increased from 4 to 5 years of age.

Preston et al (2010) distinguish three talker groups: Late talkers, on-time talkers, and early talkers. Norms provided for TVExp-Usp allows us to distinguish among five talker groups (very late talking, late talking, on-time talking, early talking, very early talking) at each of five age levels (i.e., from 1 to 5 years of age).

Table 1 summarizes the Number of pictures (from 1 to 100) that are named by children from each of the five skill levels (very late talking, late talking, on-time talking, early talking, very early talking) for each of the five age levels (1 to 5 years of age).

As summarized by the table: For one year-old children, on average: On-time talkers name 24 to 27 pictures, late talkers name 23 pictures, and very late talkers name only 22 of them; whereas early talkers name 28 pictures, and very early talkers name 29 of them. The table also indicates that for two year-old children, on average, on-time talkers name 50 to 76 pictures, late talkers name 37 to 49 pictures, and very late talkers name only 25 to 36 of them; whereas early talkers name 77 to 89 pictures, and very early talkers name 90 to 100 of them.

Results also showed that success in oral naming was a positive function of picture iconicity and word familiarity as measured by Capovilla's *Speech-Invoking Pictography* (Capovilla, Negrão, Damazio, Roberto et al, 2011).

Thus, speech production may be assessed with the LAVE inventory for 2-6 year old children, or with the TVExp-Usp test for 18 month-old children to 6 year old ones. TVExp-Usp test is superior to LAVE inventory because it provides a direct measure of speech production, and because it permits identifying five levels of talking skills at each of five age levels, from 1 to 5 years of age.

Instruments for assessing speech comprehension in Brazil

PPVT Brazilian version: Teste de Vocabulário por Imagens Peabody (TVIP)

Peabody Picture Vocabulary Test (PPVT: Dunn & Dunn, 1981) is one of the most widely used tests for assessing the development of auditory vocabulary (word comprehension) from 2 years and 6 months of age until 18 years of age. It provides an objective, fast and precise measure of auditory vocabulary development in a wide variety of areas including people, actions, qualities, body parts, time, nature, places, objects, animals, tools and instruments, and mathematical terms.

The Spanish version of PPVT is called *Test de Vocabulario en Imágenes Peabody* (Dunn, Padilla, Lugo, & Dunn, 1986a, 1986b). The Brazilian version of PPVT is called *Teste de Vocabulário por Imagens Peabody* (TVIP). It has been adapted by Capovilla and Capovilla (1997). TVIP consists of five pretest items, and 125 test items, which are organized in increasing difficulty level. Each item is made of a card with four line drawings. Each line drawing is printed in black against a white background. The examiner reads from a word list. At each word pronounced by the examiner, the child is to

match the word by selecting one out of four pictures. On each item the examinee is not required to read, speak or write anything, but only to mark (or point at) the picture that best corresponds to the word spoken by the examiner. In this multiple choice test with four alternatives, there is a 25% chance of scoring by pure guessing. Thus, word comprehension score varies from 31 points to 125 points.

TVIP validity, reliability and standardization have been established by Capovilla and Capovilla (1997) for 2-6 year old children, and by Capovilla, Nunes, Araújo, Nogueira, and Bernat (1997), Capovilla, Nunes, Nunes et al. (1997), and Capovilla, Nunes, Nogueira et al. (1997) for 6-14 year old children.

Capovilla and Capovilla (1997) evaluated 238 children from 2 to 6 years of age (103 from public schools, and 135 from private schools). ANOVA showed that TVIP scores were a positive monotonic function of examinees' school grade, $F(3, 111) = 100.2, p = .000$ and age in both years, $F(4, 110) = 56.04, p = .000$ and semesters, $F(8, 106) = 40.35, p = .000$. Based on such data, standardization tables were generated. In addition, results showed that school type (public *versus* private) also had effect: Private school examinees scored higher than did public school ones, $F(1, 145) = 15.17, p = .000$.

Regression analysis of TVIP scores on word production inventories scores revealed a modest positive correlation, $r = .36$, $F(1, 120) = 17.5$, $p = .000$. This means that children who scored higher in word production also scored higher in word comprehension.

Capovilla, Nunes et al. (1997) evaluated 687 public school students. The sample encompassed examinees from early childhood education (2-3 years of age) through preschool (or kindergarten: ages 4-5), primary school (ages 6-10 and grades 1-5), junior high school (ages 11-14 and grades 6-9), until senior high school (or secondary school: ages 15-18 and grades 10-12).

Data analysis was conducted using two types of scoring criteria: Raw scoring and Basal-ceiling scoring. Raw scores consisted simply of the number of correct responses, that is, the number of test items to which a given examinee responded properly (i.e., number of hits). Basal-ceiling scores consisted of the number of correct responses within the range between Basal level and Ceiling level, that is, excluding points pertaining to both: lower range (i.e., easy items, which are below the Basal level criteria for age and school grade), and upper range (i.e., difficult items, which are above the Ceiling level criterion). The study aimed at assessing item difficulty level so as to permit

ordering items based on that level by means of item analysis. It also aimed at establishing Basal and Ceiling parameters, so as to permit comparing the effects of the two scoring criteria (i.e., Raw scoring versus Basal-ceiling scoring) upon norms (i.e., standardization data) in such an early stage of test adaptation. Table 2 summarizes normative data obtained with both scoring criteria: Basal-ceiling scoring and Raw scoring.

Table 2. TVIP norms: Number of pictures (from 1 to 125) that are pointed by children from each of the five skill levels (very late listening, late listening, on-time listening, early listening, very early listening) for each of the nine school grade levels. Top part: Basal-ceiling scoring method; Bottom part: Raw scoring method.

School grade levels	Very late listeners	Late listeners	On-time Listeners	Early listeners	Very early listeners
	-3SD to -2SD	-2SD to -1SD	-1SD to +1SD	+1SD to +2SD	+2SD to +3SD
Basal-ceiling scoring method					
K	21 - 35	36 - 49	50 - 79	80 - 93	94 - 108
1st	37 - 49	50 - 61	62 - 87	88 - 99	100 - 111
2nd	55 - 65	66 - 76	77 - 99	100 - 110	111 - 122
3rd	55 - 66	67 - 78	79 - 104	105 - 116	117 - 125
4th	66 - 77	78 - 89	90 - 115	116 - 125	---
5th	84 - 92	93 - 101	102 - 121	122 - 125	---
6th	98 - 103	104 - 109	110 - 121	122 - 125	---
7th	109 - 112	113 - 115	116 - 123	124 - 125	---
8th	12 - 114	115 - 117	118 - 123	124 - 125	---
Raw scoring method					
K	44 - 54	55 - 64	65 - 85	86 - 95	96 - 105
1st	55 - 62	63 - 71	72 - 90	91 - 99	100 - 108
2nd	68 - 75	76 - 83	84 - 100	101 - 108	109 - 116
3rd	70 - 78	79 - 86	87 - 103	102 - 112	113 - 120
4th	81 - 87	88 - 95	96 - 110	111 - 117	118 - 124
5th	89 - 94	95 - 101	102 - 114	115 - 121	122 - 125
6th	99 - 102	103 - 107	108 - 116	117 - 121	122 - 125
7th	105 - 107	108 - 111	112 - 120	121 - 124	125
8th	107 - 109	110 - 112	113 - 119	120 - 121	122 - 125

Data analysis of Basal-ceiling scores

In terms of Basal-ceiling scores (i.e., data obtained with the Basal-ceiling scoring criterion), ANOVA of TVIP scores as function of school grade level revealed a significant effect of school grade level: $F(8, 671) = 333.81, p < .0001$. Bonferroni pair wise analysis revealed that TVIP scores increased significantly ($p < .0001$) from 1st grade (64.1 points) to 2nd grade (73.9 points), and from then to 3rd grade (87.9 points). They also revealed that TVIP scores increased significantly from 4th grade (91.1 points) to 5th grade (102.3 points), and from then to 6th grade (111.7 points).

Data analysis of Raw scores

In terms of Raw scores (i.e., data obtained with the Raw scoring criterion), ANOVA of TVIP scores as function of school grade level revealed significant effect of school grade level: $F(8, 671) = 372.78, p < .0001$. Bonferroni pair wise analysis revealed that TVIP scores increased significantly ($p < .0001$) from 1st grade (74.6 points) to 2nd grade (81.0 points), and from then to 3rd grade (92.3 points). They also revealed that TVIP scores increased significantly ($p < .0001$) from 4th grade (91.1 points) to 5th grade (102.8 points), from then to 6th grade (108.1 points), from then to 7th grade

(112.0 points), and from then to 8th grade (116.4 points).

Comparison between the two scoring criteria

Comparing data obtained from the two scoring criteria, one can infer that Raw scoring criterion tends to overestimate the capabilities (scores) of younger examinees, whereas the Basal-ceiling scoring criterion tends to overestimate the capabilities of older examinees

The reason why the Raw scoring criterion tends to overestimate scores of younger examinees is that such a criterion assumes that all hits that occur on the upper difficulty level are legitimate hits; and the proportion of casual hits to legitimate hits increases as the difficulty level increases and the age level decreases. That is: The proportion of casual hits to legitimate hits is directly proportional to the item difficulty level, and inversely proportional to the examinees' age and school grade level. In sum: The younger the examinee, and the harder the item, the greater the proportion of casual hits to legitimate hits.

The reason why the Basal-ceiling scoring criterion tends to overestimate scores of older examinees is that such a criterion assumes that all responses that would occur in the lower difficulty level (below the basal level) would be hits (if

they were to be tested), and thus it grants these items full score. And the proportion of casual hits to legitimate hits increases as the difficulty level increases and the age level decreases. Such a scoring criterion presupposes that all responses to the items regarded as being too easy would constitute hits if they happened to be tested. Thus, the scoring criterion attributes full score to the items pertaining to the lower range of difficulty level. The benefit (the bonus/onus ratio) that results from such an assumption is directly proportional to the age of the examinee.

Therefore, one might suppose that the most appropriate estimate of the real score would be obtained by adopting the Basal-ceiling scoring criterion for 1st to 3rd graders, and to adopt the Raw scoring criterion for 5th to 8th graders. Either scoring criterion might be used for 4th graders, since they belong to the middle of the scale, and thus, the risk of underestimation and overestimation would be minimal. In addition, one investigator might choose to adopt Basal-ceiling scoring criterion in order to benefit from the interruption criterion and to prevent children from experiencing repeated failure when tested to exhaustion.

TVIP internal consistency reliability was assessed via split half coefficient, which was equal to .914, $p < .0001$. Spearman-Brown correction was

equal to .955. Kuder Richardson Coefficient of reliability (K-R 20) was equal to .951 for Raw scoring and to .977 for Basal-ceiling scoring.

The University of Sao Paulo Receptive Vocabulary Test: Teste de Vocabulário por Figuras Usp (TVF-Usp)

The original version of the *University of Sao Paulo Receptive Vocabulary Test* (TVF-Usp) was published originally by Capovilla and Prudencio (2006). It permits assessing the development of auditory vocabulary (word comprehension) from 2 to 14 years of age. It provides a royalty-free, objective, fast and precise measure of auditory vocabulary development. TVF-Usp consists of five pretest items, and 139 test items, which are organized in increasing difficulty. Each item is made of a card with four alternative line drawings. Each line drawing is printed in black against a white background. Figure 1, middle row, shows one of 139 test items of TVF-Usp.

On each item the examinee is not required to read, speak or write anything, but only to mark (or point at) the picture that better corresponds to the word spoken by the examiner. In this multiple choice test with four alternatives, there is a 25% chance of scoring by pure guessing. Thus, word comprehension score varies from 35 points to 139 points.

TVF-Usp validity, reliability and standardization have been established in a series of six studies summarized by Capovilla (2011). In those studies, TVF-Usp scores have been compared with performance in a dozen of standardized tests covering skills such as: phonological awareness, auditory comprehension, reading comprehension, central auditory processing, spelling under dictation, visual speechreading, phonological memory, and intelligence. Norms have been established for six different socio-economic status levels. One of those studies was conducted with 629 extremely poor children who lived in shantytowns.

The sample included 222 1st graders (aged 7 years and 11 months), 181 2nd graders (aged 8 years and 10 months), and 226 3rd graders (aged 9 years and 10 months). Children were evaluated collectively in their classroom in two 45 min sessions. Scores varied from 23 to 131, with a mean of 101.52 and standard deviation of 14.8. Regression analysis showed that TVF-Usp scores were significantly correlated with school grade levels: $F(1, 628) = 179.35, p < .000, N = 629, r = .47, r^2 = .22$, standard error of the estimate = 13.06. ANOVA of auditory comprehension as a function of school grade level revealed significant effect: TVF-Usp scores were a positive monotonic function of examinees' school

grade level, $F(2, 628) = 89.57, p < .000, N = 629$, partial eta squared $\eta^2 = .22$. Bonferroni pair wise comparison revealed significant increase from 1st grade (93.9 points) to 2nd grade (101.3 points), and from then to 3rd grade (109.8 points).

Regression analysis showed that TVF-Usp scores were significantly correlated with age level: $F(1, 628) = 94.97, p < .000, N = 629, r = .36, r^2 = .13$, standard error of the estimate = 13.08. ANOVA of auditory comprehension as a function of age level revealed significant effect: TVF-Usp scores were a positive monotonic function of examinees' age level, $F(5, 623) = 21.62, p < .000, N = 629$, partial eta squared $\eta^2 = .15$. Bonferroni pair wise comparison revealed significant increase from 6 years (90.9 points) and 7 years (94.8 points) to 8 years (100.5 points) of age, and from 8 years (100.5 points) to 9 years (108.0 points) of age.

Auditory Vocabulary Test: Teste de Vocabulário Auditivo (TVAud): unabridged (TVAud-107) and abridged versions (TVAud-33)

The *Auditory Vocabulary Test* (TVAud) is a standardized test for assessing word comprehension in children from 18 months of age until 6 years of age. There are two unabridged forms with 107 items each (forms A107 and B107) and

two abridged forms with 33 items each (forms A33 and B33). Each item is made of five pictures: one target and four distracters. Figure 1, bottom row, illustrates one of the test items. Each item begins with the examiner speaking a word. The examinee's task is to point to the target picture that corresponds to the word that was spoken by the examiner. Examinees from 18 to 24 years of age are evaluated individually, and are required to point to target pictures. Examinees from 2 to 6 years of age are evaluated in small groups and are required to mark target pictures in the card. There are five items per card. To the left side of each item there is an index picture, which serves to help children to focus on the item to be responded to. There are five types of index pictures: arrow, triangle, star, heart, and square. At the beginning of each item, the examiner tells the examinees to search for the appropriate index picture to the left (which may be either an arrow, or a triangle, or a star, or a heart, or even a square), and then to focus on the sequence of pictures that begins with that index picture. And finally to point to the target picture that corresponds to the word that was spoken.

TVAud unabridged version (TVAud-107), forms A and B

Test administration of either one of the two unabridged forms (A107 or B107)

takes 30 min for 4, 5 and 6 year-old children. It takes 60 min for 3 year-old children, and 120 min (divided in several sessions) for 2 year-old children. A107 and B107 are administered individually for 2 year-old children and collectively in small groups, of 10-15 children each, for 3 to 6 year-old children.

Capovilla, Negrão, and Damazio (2011) evaluated 398 2-5 year-old children. The study provided data on the validity and reliability of the two forms of the unabridged version (Forms A107 and B107). Items analysis permitted the exclusion of items with low item-test correlation and reordering according to increasing difficulty. Furthermore, it permitted the construction of the abridged forms containing only 33 items each (A33 and B33 forms).

With respect to A107, regression analysis showed that A107 test scores increased significantly with age: $F(1, 397) = 83.13, p < .000, N = 398, r = .42, r^2 = .18$, standard error of the estimate = 3.99. ANOVA of auditory comprehension as a function of age levels revealed significant effect: A107 scores were a positive monotonic function of examinees' age level, $F(3, 398) = 31.77, p < .000$, partial eta squared $\eta^2 = .20$. Bonferroni pair wise comparison revealed significant increase in receptive vocabulary scores from 2 years of age (94.9 points) to 3 years of age

(101.22 points), from then to 4 years of age (102.62 points), and from then to 5 years of age (104.96 points). A107 was standardized from 2 to 5 years of age.

With respect to B107, regression analysis showed that B107 test scores also increased significantly with age: $F(1, 397) = 13.83, p < .000, N = 398, r = .18, r^2 = .03$. ANOVA of auditory comprehension as a function of age level revealed significant effect: B107 scores were a positive monotonic function of examinees' age level, $F(3, 398) = 5,187, p < .000$, partial eta squared $\eta^2 = .04$. Fisher LSD pair wise comparison revealed that Receptive Vocabulary scores increased significantly from age 2 (99.2 points) to age 3 (101.2 points), and from ages 3 (101.2 points) and 4 (101.8 points) to age 5 (104.4 points). B107 was standardized from 2 to 5 years of age. Bonferroni pair wise comparison revealed that Receptive Vocabulary scores increased significantly from 4 to 5 years of age.

Considering the 396 children who were exposed to both A107 and B107 forms, regression analysis showed significant correlation between A107 and B107 forms, $F(1, 394) = 200.66, p < .000, N = 396, r = .58, r^2 = .34$. Thus, B107 form can be used as a secondary control form for heuristic research purposes in children from 2 to 5 years of age.

TVAud abridged version (TVAud-33) forms A and B

The *Auditory Vocabulary Test* (TVAud) abridged version, forms (A33 and B33), derive from the two forms of the original unabridged version (Forms A107 and B107) after Capovilla, Negrão, and Damazio (2011) study with 2-6 year old children. Forms A33 and B33 are made of 7 A4-size boards. Each board contains five items. Each item is made of a sequence of five pictures, of which one is the target and the others are distracters. The examinee's task consists of choosing the picture that best corresponds to the word spoken by the examiner. One and two-year old children point at the picture, whereas 3-6 year old children mark the picture.

Test administration of either one of the two abridged forms (A33 or B33) takes 15 min for 4, 5 and 6 year old children. It takes 30 min for 2-3 year-old children, and 60 min (divided in several sessions) for 1.5 year-old children. A33 and B33 are administered individually for 1.5 and 2 year old children and collectively in small groups, of 10-15 children each, for 3 to 6 year old children.

Capovilla, Negrão, and Damazio (2011) conducted a study with 177 1-6 year old children. Number of children per age group was as follows: 10 children aged 1 year ($M = 1y8m$), 28 2 years ($M = 2y7m$), 26 3 years ($M = 3y6m$), 38 4 years

($M = 4y6m$), 34 5 years ($M = 5y7m$), and 41 5 years ($M = 6y5m$).

With respect to A33 vocabulary scores, ANOVA revealed a significant increase with age, $F(5, 171) = 79.64$, $p < .000$, $N = 177$, $r = .84$, $r^2 = .70$, partial eta squared $\eta^2 = .70$. Therefore, increase in age accounted for 70% of the total variability in the vocabulary scores. Fisher LSD pair wise analysis revealed that auditory vocabulary increased from 1 to 2 years of age, from 2 to 3 years of age, from 3 to 4 years of age, and from 4 to 5 years of age. Bonferroni pair wise analysis revealed that auditory vocabulary increased from 1 to 2 years of age, from 2 to 3 years of age, and from 4 to 5 years of age.

With respect to B33 vocabulary scores, ANOVA revealed a significant increase with age, $F(5, 171) = 81.22$, $p < .000$, $N = 177$, $r = .84$, $r^2 = .70$, partial eta squared $\eta^2 = .70$. Fisher LSD pair wise analysis revealed that auditory vocabulary increased from 1 to 2 years of age, from 2 to 4 years of age, from 3 and 4 to 5 years of age. Bonferroni pair wise analysis revealed that auditory vocabulary increased from 1 to 2 years of age, and from 2, 3 and 4 to 5 and 6 years of age.

Considering the 177 children who were exposed to both A33 and B33 forms, regression analysis showed significant correlation between A33 and B33 forms, F

(1, 175) = 804.52, $p < .000$, $N = 177$, $r = .91$, $r^2 = .82$. Thus, B33 form can be used as a secondary control form for heuristic research purposes in children from 1 to 5 years of age.

Thus, speech comprehension may be evaluated with TVIP (the Brazilian version of PPVT) and TVF-Usp, both of them for 2 to 14 year-old children; or with TVAud100 and TVAud33, both of them for 1.5 to 6 year-old ones. TVF-Usp is superior to TVIP not only because it a royalty free test, but also because its pictures have been selected from a pool of over 10,000 pictures after careful analyses of their iconicity and name agreement properties (Capovilla, Negrão, Damazio et al., 2011). In addition to that, its criterion related validity has been strongly established in Brazil with respect to a number of different measures of linguistic (e.g., reading, spelling), and metalinguistic competences (e.g., phonological awareness, morpho-syntactic awareness), as well as skills pertaining to central auditory processing, memory and intelligence, among others (Capovilla, 2011).

TVAud107 and TVAud33 are superior to both TVF-Usp and TVIP because they permit earlier assessment (from 18 months on, rather than only from 24 months on) and because they both have alternate forms (TVAud-A107 and

TVAud-B107; as well as TVAud-A33 and TVAud-B33). TVAud33 is superior to TVAud107 because it provides a quicker assessment (with 33 items instead of 107 ones) while preserving its sensitivity and precision.

Instrument for assessing sign comprehension in Brazil

Assessing Brazilian Sign Language: Libras Sign Receptive Test: *Teste de Vocabulário Receptivo de Sinais de Libras* (TVRSL)

Brazilian sign language (Libras) is the native language of the Brazilian deaf population. The lexicon of Libras has been documented in three encyclopedic dictionaries: *Deit-Libras* (Capovilla & Raphael, 2001a, 2001b), *Novo Deit-Libras* (Capovilla, Raphael, & Mauricio, 2012a, 2012b), and *Dic Brasil* (Capovilla, Raphael, Temoteo, & Martins, in print a, in print b, in print c), as well as in the *Brazilian Sign Language Encyclopedia* (Capovilla & Raphael, 2004a, 2004b, 2005a, 2005b, 2005c).

Brazilian education is in need of standardized instruments for assessing language development of the deaf school population. In order to start filling that gap, Capovilla and Raphael (2004b) have created the *Libras Sign Receptive Test* (TVRSL-139), to assess Libras sign comprehension in Brazilian deaf students.

TVRSL is made of 139 items. Each item is made of a sequence of four pictures. On each item, the examiner performs one sign in front of the class. The examinees' task is to choose the target picture that corresponds to the meaning of the sign, while ignoring the other three distracter pictures. The target picture is to be marked by pencil. Using the abridged version with only 66 items (TVRSL-66), the authors have recently examined 4011 deaf students from 2nd to 8th school grade levels, aging 6 to 41 years. The majority were pre-lingual deaf. TVRSL-66 was administered collectively in the classroom with live signing by a fluent native signer. ANOVA of sign comprehension score as a function of school grade level revealed significant effect of school grade level, $F(6, 4011) = 138.85$, $p < .000$, $N = 4011$, $r^2 = .17$, partial eta squared $\eta^2 = .17$. Bonferroni pair wise comparison revealed Libras sign comprehension scores increased from 2nd grade (24,09 points) to 3rd grade (27,77 points), from 3rd to 4th grade (31,08 points), from 4th to 5th (34,77 points), from 5th to 6th (36,95 points), from 6th to 7th (40,50 points), and from 7th to 8th (44,02 points). TVRSL-66 was standardized for 2nd grade to 8th grade students. Norms provided for TVRSL-66 permit distinguishing among five sign recipient groups (very late receivers, late receivers, on-time receivers, early receivers, very

early receivers) at each of seven school grade levels (i.e., from 2nd grade to 8th grade). Table 3 summarizes the Number of signs (from 1 to 66) that are received by deaf children from each of the five sign comprehension skill levels (very late receiving, late receiving, on-time receiving, early receiving, very early receiving) for each of seven school grade levels (i.e., from 2nd grade to 8th grade) in TVRSL-66.

Table 3. Number of signs (from 1 to 66) that are received and comprehended by deaf children from each of the five sign comprehension skill levels (very late receiving, late receiving, on-time receiving, early receiving, very early receiving) for each of seven school grade levels (i.e., from 2nd grade to 8th grade).

School grade levels	Very late receivers -3SD to -2SD	Late receivers -2SD to -1SD	On-time receivers -1SD to +1SD	Early receivers +1SD to +2SD	Very early receivers +2SD to +3SD
2 nd	0-2	3-13	14-35	36-45	46 - 56
3 rd	0	1-14	15-41	42-54	55 - 66
4 th	0-5	6-17	18-44	45-56	57 - 66
5 th	0-9	10-23	24-49	50-62	63 - 66
6 th	0-11	12-23	24-50	51-62	63 - 66
7 th	0-8	9-24	25-56	57-72	---
8 th	0-12	13-27	28-60	61-75	---

Conclusion

The standardized tools presented in this paper permit assessing the development of both receptive vocabulary and expressive vocabulary in Brazilian Portuguese and Brazilian Sign Language in both hearing and deaf Brazilian children from 18 months to 14 years of age.

Speech production may be assessed with the LAVE inventory (for 2-6 years of age), or with the TVExp-Usp test (for 18-72 months of age). TVExp-Usp is superior to LAVE because it provides a direct measure of speech production, and because it permits identifying five levels of talking skills from one year to the next (for 1-5 years of age).

Speech comprehension may be evaluated with TVIP and TVF-Usp (both for 2-14 years of age); or with TVAud100 and TVAud33, (both for 1.5 to 6 years of age). TVF-Usp is superior to TVIP because: 1) it a royalty free test; 2) its pictures have been selected from a pool of over 10,000 pictures after establishing their iconicity and name agreement properties; 3) its criterion related validity has been established in Brazil with respect to several measures of linguistic and metalinguistic skills. TVAud107 and TVAud33 are superior to both TVF-Usp and TVIP because they: 1) permit earlier assessment (from 18 months on) and 2) have alternate forms. TVAud33 is superior

to TVAud107 because it provides a quicker assessment.

The normative data provided herein allow early identification of children in need for early preventive intervention, as well as early comparisons of treatment effectiveness. Norms provided herein for TVExp-Usp permit distinguishing among five talker groups (very late talkers, late talkers, on-time talkers, early talkers, very early talkers) at each of five age levels (i.e., from 1 to 5 years of age). By the same token, norms provided herein for TVIP permit distinguishing among five listener groups (very late listeners, late listeners, on-time listeners, early listeners, very early listeners) at each of nine school

grade levels (i.e., from Kindergarten to 8th grade).

Libras sign comprehension in deaf children is a strong predictor of school performance, and may be evaluated with TVRSL (for 2-14 years of age). Norms provided for TVRSL-66 permit distinguishing among five sign recipient groups (very late receivers, late receivers, on-time receivers, early receivers, very early receivers) at each of seven school grade levels (i.e., from 2nd grade to 8th grade). The instruments described herein, are important tools for practice, education, and research in expressive and receptive language development in Brazil.

References

Burgemeister, B., Blum, L., & Lorge, I. (1971). *Columbia Mental Maturity Scale*. New York, NY: Harcourt, Brace & Ovanovich.

Capovilla, F. C. (2011). *Teste de Vocabulário por Figuras USP: normarizado para avaliar a compreensão auditiva dos 7 aos 10 anos*. São Paulo, SP: Memnon Edições Científicas. (530 pp. ISBN: 978-85-794-016-5).

Capovilla, F. C., & Capovilla, A. G. S. (1997). Desenvolvimento linguístico da criança dos dois aos seis anos: Tradução e estandardização do Peabody Picture Vocabulary Test de Dunn, & Dunn, e da Language Development Survey de Rescorla. *Ciência Cognitiva: Teoria, Pesquisa e Aplicação*, 1(1), 353-380. (ISSN: 1415-1472).

Capovilla, F. C., & Damazio, M. (2011). *Teste de Vocabulário Expressivo*.

- Em F. C. Capovilla, V. B. Negrão, & M. Damazio (2011). *Teste de Vocabulário Auditivo e Teste de Vocabulário Expressivo*. São Paulo, SP: Memnon Edições Científicas Ltda (ISBN: 978-85-7954-017-2).
- Capovilla, F. C., Negrão, V. B., & Damazio, M. (2011). *Teste de Vocabulário Auditivo e Teste de Vocabulário Expressivo*. São Paulo, SP: Memnon Edições Científicas Ltda (ISBN: 978-85-7954-017-2).
- Capovilla, F. C., Negrão, V. B., Damazio, M., Roberto, M. R., Sousa-Sousa, C. C., & Sousa, A. V. (2011). *Pictografia evocadora da fala: CD-Rom e Manual*. São Paulo, SP: Casa do Psicólogo.
- Capovilla, F. C., Nunes, L. R. O. P., Araújo, I., Nogueira, D., & Bernat, A. B. (1997). Versão brasileira do Teste de Vocabulário Por Imagens Peabody. *Revista Distúrbios da Comunicação*, 8(2), 151-162.
- Capovilla, F. C., Nunes, L. R. O. P., Nogueira, D., Nunes, D., Araújo, I., Bernat, A. B., & Capovilla, A. G. S. (1997). Análise da validade concorrente do Teste de Vocabulário por Imagens Peabody por comparação com o desempenho escolar de pré-escola a 8^a série: Amostra fluminense. *Ciência Cognitiva: Teoria, Pesquisa e Aplicação*, 1(2), 533-560. (ISSN: 1415-1472).
- Capovilla, F. C., & Prudencio, E. R. (2006). Teste de Vocabulário Auditivo por Figuras: Normatização e validação preliminares. *Avaliação Psicológica*, 5, 189-203.
- Capovilla, F. C., & Raphael, W. D. (2001a). Índice semântico dos sinais de Libras. Em F. C. Capovilla, & W. D. Raphael (Orgs.), *Dicionário enciclopédico ilustrado trilingue da Língua de Sinais Brasileira*. Vol. II: Sinais de M a Z (2^a ed., Vol. 2, pp. 1431-1434). São Paulo, SP: Edusp e

- Imprensa Oficial do Estado de São Paulo. (ISBN: 85-314-0669-2).
- Capovilla, F. C., & Raphael, W. D. (2001b). Prefácio. Em F. C. Capovilla, & W. D. Raphael (Orgs.), *Dicionário enciclopédico ilustrado trilíngue da Língua de Sinais Brasileira*. Vol. I: Sinais de A a L (2a. ed., Vol. 1, pp. 29-36). São Paulo, SP: Edusp e Imprensa Oficial do Estado de São Paulo. (ISBN: 85-314-0668-4).
- Capovilla, F. C., & Raphael, W. D. (2004a). *Enciclopédia da Língua de Sinais Brasileira: O mundo do surdo em Libras*, Vol. 1: Sinais de Libras e o universo da educação, e Como avaliar o desenvolvimento da competência de leitura de palavras (processos de reconhecimento e decodificação) em escolares surdos do Ensino Fundamental ao Médio (pp. 1-680). São Paulo, SP: Edusp, Imprensa Oficial do Estado de São Paulo, Fundação Vitae, Capes, CNPq, e Fapesp. (ISBN: 85-314-0849-0, ISBN: 85-7060-276-6).
- Capovilla, F. C., & Raphael, W. D. (2005a). *Enciclopédia da Língua de Sinais Brasileira: O mundo do surdo em Libras*, Vol. 3: Sinais de Libras e a vida em família, relações familiares e casa, e Como avaliar o desenvolvimento da competência de leitura de sentenças (processamento
- CNPq, e Fapesp. (ISBN: 85-314-0826-1, ISBN: 85-7060-269-3).
- Capovilla, F. C., & Raphael, W. D. (2004b). *Enciclopédia da Língua de Sinais Brasileira: O mundo do surdo em Libras*, Vol. 2: Sinais de Libras e o universo das artes e cultura, esportes, e lazer, e Como avaliar o desenvolvimento da compreensão de sinais (vocabulário em Libras) de escolares surdos de 1a. a 8a. série do Ensino Fundamental (pp. 1-827). São Paulo, SP: Edusp, Imprensa Oficial do Estado de São Paulo, Fundação Vitae, Capes, CNPq, e Fapesp. (ISBN: 85-314-0849-0, ISBN: 85-7060-276-6).

- sintático e semântico) de escolares surdos do Ensino Fundamental ao Médio (pp. 1–857). São Paulo, SP: Edusp, Imprensa Oficial do Estado de São Paulo, Fundação Vitae, Capes, CNPq, e Fapesp. (ISBN: 85-314-0855-5).
- Capovilla, F. C., & Raphael, W. D. (2005b). *Enciclopédia da Língua de Sinais Brasileira: O mundo do surdo em Libras*, Vol. 4: Sinais de Libras e o universo da comunicação, eventos e religião, e Como avaliar a competência de leitura (processamento quirêmico e ortográfico) de escolares surdos do Ensino Fundamental ao Médio (pp. 1–1010). São Paulo, SP: Edusp, Imprensa Oficial do Estado de São Paulo, Fundação Vitae, Capes, CNPq, e Fapesp. (ISBN: 85-314-0870-9).
- Capovilla, F. C., & Raphael, W. D. (2005c). *Enciclopédia da Língua de Sinais Brasileira: O mundo do surdo em Libras*, Vol. 8: Sinais de Libras e o mundo das palavras de função gramatical, e Como acompanhar o desenvolvimento da competência de leitura (processos quirênicos, semânticos e ortográficos) de escolares surdos do Ensino Fundamental ao Médio. São Paulo, SP: Edusp, Fundação Vitae, Capes, CNPq, e Fapesp. (ISBN: 85-314-0902-0).
- Capovilla, F. C., Raphael, W. D., & Mauricio, A. C. (2012a) (Orgs.), *Novo Deit-Libras: Dicionário enciclopédico ilustrado trilíngue da Língua de Sinais Brasileira (Libras) baseado em Linguística e Neurociências Cognitivas*, 3ª edição, Volume 1. São Paulo, SP: Edusp. (ISBN: 85-314-1433-6).
- Capovilla, F. C., Raphael, W. D., & Mauricio, A. C. (2012b) (Orgs.), *Novo Deit-Libras: Dicionário enciclopédico ilustrado trilíngue da Língua de Sinais Brasileira (Libras)*

- baseado em Linguística e Neurociências Cognitivas*, 3ª edição, Volume 2. São Paulo, SP: Edusp. (ISBN: 85-314-1434-6).
- Capovilla, F. C., Raphael, W. D., Temoteo, J., & Martins, A. C. (in print a). *Dic Brasil: Dicionário da língua de sinais brasileira: A Libras em suas mãos*. Vol. I: Sinais de A a D. São Paulo, SP: Edusp e Imprensa Oficial do Estado de São Paulo.
- Capovilla, F. C., Raphael, W. D., Temoteo, J., & Martins, A. C. (in print b). *Dic Brasil: Dicionário da língua de sinais brasileira: A Libras em suas mãos*. Vol. II: Sinais de E a O. São Paulo, SP: Edusp e Imprensa Oficial do Estado de São Paulo.
- Capovilla, F. C., Raphael, W. D., Temoteo, J., & Martins, A. C. (in print c). *Dic Brasil: Dicionário da língua de sinais brasileira: A Libras em suas mãos*. Vol. III: Sinais de P a Z. São Paulo, SP: Edusp e Imprensa Oficial do Estado de São Paulo.
- Cronbach, L. (1996). *Fundamentos de testagem psicológica* (5ª ed.). Porto Alegre, RS: Artes Médicas.
- Dunn, L. M., & Dunn, D. M. (1997). *PPVT-III: Peabody Picture Vocabulary Test - III*. 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. (1986a). *Manual del examinador para el Test de Vocabulario en Imágenes Peabody*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., Padilla, E. R., Lugo, D. E., & Dunn, L. M. (1986b). *Test de Vocabulario en Imágenes Peabody: Adaptación Hispanoamericana*. Circle Pines, MN: American Guidance Service.
- Ferracini, F., Capovilla, A. G. S., Dias, N. M., & Capovilla, F. C. (2006).

- Avaliação de vocabulário expressivo e receptivo na educação infantil. *Psicopedagogia: Revista da Associação Brasileira de Psicopedagogia*, 23(71), 124-133. (ISSN: 0103-8486).
- Pedromônico, M. R. M., Affonso, L. A., & Sãnudo, A. (2002). Vocabulário expressivo de uma amostra de crianças entre 22 e 36 meses: estudo exploratório. *Revista Brasileira de Crescimento e Desenvolvimento Humano*, 12(2),13-22.
- Preston, J. L., Frost, S. J., Mencl, W. E., Fulbright, R. K., Landi, N., Grigorenko, E., Jacobsen, L., & Pugh, K. R. (2010). Early and late talkers: school-age language, literacy and neurolinguistic differences. *Brain*, 133, 2185-2195. doi:10.1093/brain/awq163.
- Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, 54(4), 587-599.