

Development and validation of lists of disyllabic words for speech audiometry testing

Elaboração e validação de listas de dissílabos gravados para teste logoaudiométrico

Tais Regina Hennig¹, Ana Valéria de Almeida Vaucher¹, Maristela Julio Costa²

ABSTRACT

Purpose: To compose a bank of disyllabic words to develop equivalent disyllabic lists, perform content validation, obtain evidence of reliability and digitally record these lists to determine the Speech Recognition Percentage Index (SRPI) in order to complement the set of materials available for this evaluation. **Methods:** We used disyllabic, paroxytone nouns, which were submitted to content validation, which included assessment of familiarity, appropriateness and auditory recognition by expert and non-expert raters. Lists of disyllabic words (with 25 words each) were developed from the words selected after content validation, and the equivalence search of these lists was carried out to collect evidence of reliability for the proposed new test instrument. **Results:** The first version of the word bank was composed of 442 disyllables; 198 of them were considered to be familiar by most raters, and 176 were deemed as appropriate; after auditory recognition, 172 words were kept in the word bank, distributed into six lists, with 25 words in each one. Among these lists, only one differed from the others while the other five were considered to be equivalent, and were named LD-A, LD-B, LD-C, LD-D and LD-E, and recorded in digital format onto a Compact Disc. **Conclusion:** Five lists of disyllabic words were considered as equivalent, named LD-A, LD-B, LD-C, LD-D and LD-E lists. They were digitally recorded and made available with satisfactory evidence of validity and reliability, to complement the set of available speech materials for SRPI assessment.

Keywords: Hearing; Audiometry speech; Speech perception; Speech discrimination tests; Psychometry

RESUMO

Objetivo: Compor um banco de palavras dissilábicas para elaborar, realizar a validação de conteúdo, obter evidências de fidedignidade e gravar digitalmente listas de dissílabos equivalentes, para a realização do Índice Percentual de Reconhecimento de Fala (IPRF), a fim de complementar a bateria de materiais de fala disponíveis para essa avaliação. **Métodos:** Foram selecionados vocábulos dissilábicos, paroxítonos, substantivos, os quais foram submetidos ao processo de validação de conteúdo, que abrangeu o julgamento quanto à familiaridade, adequação e reconhecimento auditivo, por juízes especialistas e não especialistas. Foram elaboradas listas de dissílabos, com 25 palavras em cada uma, a partir dos vocábulos resultantes da validação de conteúdo, e realizada a pesquisa de equivalência dessas listas, a fim de obter evidências de fidedignidade para o novo instrumento de teste proposto. **Resultados:** A primeira versão do banco de palavras foi composta por 442 dissílabos. Destes, 198 foram considerados familiares pela maioria dos juízes, sendo que 176 foram julgados como adequados. Após o reconhecimento auditivo, foram mantidos, no banco de palavras, 172 vocábulos, distribuídos em seis listas, com 25 palavras em cada uma. Dentre estas listas, apenas uma diferiu das demais e cinco foram consideradas equivalentes, denominadas LD-A, LD-B, LD-C, LD-D e LD-E, gravadas em formato digital em *Compact Disc*. **Conclusão:** Cinco listas de dissílabos elaboradas foram consideradas equivalentes, nomeadas de listas LD-A, LD-B, LD-C, LD-D e LD-E, disponibilizadas em gravação digital, com evidências satisfatórias de validade e confiabilidade, para complementar a bateria de materiais de fala disponíveis para a realização do IPRF.

Palavras-chave: Audição; Audiometria da fala; Percepção da fala; Testes de discriminação da fala; Psicometria

Study carried out at Curso de Doutorado do Programa de Pós-graduação em Distúrbios da Comunicação Humana – PPGDCH, Universidade Federal de Santa Maria – UFSM – Santa Maria (RS), Brasil.

¹Programa de Pós-graduação (Doutorado) em Distúrbios da Comunicação Humana, Universidade Federal de Santa Maria – UFSM – Santa Maria (RS), Brasil.

²Departamento de Fonoaudiologia, Universidade Federal de Santa Maria – UFSM – Santa Maria (RS), Brasil.

Conflict of interests: We declare that all authors participated sufficiently in the work, to make public its responsibility on the content and that there was no conflict of interest as to authorization for their reproduction.

Authors' contribution: TRH: analysis and interpretation of data, drafting and revising the article, final approval of the version to be published; AVAV: conception and design of the study, data collection, analysis and interpretation of data, drafting and revising the article, final approval of the version to be published; MJC: conception and design of the study, analysis and interpretation of data, drafting and revising the article, final approval of the version to be published.

Funding: Grant provided by the Coordination of Improvement of Higher Education Personnel (CAPES).

Corresponding author: Tais Regina Hennig. E-mail: tha.hennig@gmail.com

Received: August 09, 2017; **Accepted:** August 22, 2018

INTRODUCTION

The most important purpose of conventional audiological evaluation is to assess speech recognition ability. Such performance is evaluated by means of threshold or sensitivity tests and suprathreshold or accuracy tests, which are part of speech audiometry testing⁽¹⁾.

Among suprathreshold or accuracy tests, the Speech Recognition (SR) represents the percentage of correct answers in a specific speech material, at an intensity that permits the best possible performance of a particular individual⁽²⁾. The level of stimulus presentation can vary between 20 and 60 dBSL, while 40 dBSL is the most frequent level^(2,3).

To determine SRPI, monosyllabic words are generally used because they are short and hence have few redundancies. Thus, to recognize them properly, individuals have to listen to all their elements⁽⁴⁾.

In Brazil, the most commonly used clinical technique for this procedure is the presentation of a list of 25 monosyllabic words, in which each item in the list represents 4% of speech recognition out of the total score⁽⁵⁾.

In addition to this recommendation, for those individuals who have difficulty in speech recognition with monosyllables, an alternative is to perform SRPI testing with disyllabic words with a view to analyzing their capacity to recognize the items with an increased number of semantic and linguistic cues^(3,6).

In this context, better auditory performance was found in recognition tasks applied to normal-hearing subjects using disyllabic stimuli, in comparison to monosyllabic words, not only with meaningful words but also pseudowords⁽³⁾. Although individuals with hearing loss present variability of answers in SRPI, the result of such assessment can aid topodiagnosis⁽⁶⁾.

As long as enough audibility is ensured, individuals with conductive deafness may show similar performance to normal hearing subjects⁽⁷⁾, while reduction of speech recognition ability in patients with sensorineural loss is proportional to their hearing loss⁽⁸⁾. When their performance is worse than expected for their degree of hearing loss, they may be affected by retrocochlear disorder⁽⁹⁾.

The selected speech material can be presented via live voice or by a recording. In the audiological routine, live voice is the most common form because it is more flexible and allows faster assessment. Also, individuals feel less tired after task application⁽⁶⁾. By contrast, a recorded material allows lesser variability of the examiner's voice characteristics^(5,6), hence it increases the reliability and validity of the test⁽¹⁰⁾.

Different materials have been developed to assess SRPI in Brazil^(3,6,11-13); some were designed for live voice presentation while others through recordings. However, there is little or scarce information available on the psychometric characteristics of these materials and on data validity and reliability⁽⁵⁾. Therefore, this study is an attempt to start filling this gap.

In this context, the aim of this work was to compose a database of disyllabic words in order to prepare a list of equivalent disyllabic words, validate their content, collect evidence of reliability, and digitally record lists with these equivalent disyllabic words to determine the Speech Recognition (SR), with a view to complementing the set of speech materials available for this type of evaluation.

METHODS

This is a cross-sectional quantitative study. This research met the ethical standards established for human research, in accordance with the Human Research Guidelines and Regulatory Standards (Resolution 466/2012 of the National Health Council), and it was duly approved by the Ethics in Human Research Committee - CEP - UFSM (13932513.1.0000.5346).

For the sake of clarity, the method for development of new speech material to determine SRPI will be described in steps.

1st step: word choice

Words were first extracted from a book⁽¹⁴⁾ which contains examples of Portuguese words with different phonemes and syllabic structures, and also from newspapers and magazines with nationwide circulation. Words were carefully selected to include common and familiar items used in various regions of Brazil so as to avoid the influence of regional usage.

Word choice covered paroxytone disyllabic nouns with the most frequent syllable structures of Portuguese: CV CV (consonant-vowel + consonant-vowel, e.g., *boca - mouth*), CVC CV (consonant-vowel-consonant + consonant-vowel, e.g., *testa - forehead*), CCV CV (consonant-consonant-vowel + consonant-vowel, e.g., *bruxa - witch*), CVCCV (consonant-vowel + consonant-consonant-vowel, e.g., *cobra - snake*)⁽¹⁵⁾.

Proper nouns (e.g., names of people or cities), plural words and pseudowords were excluded. Also, the words present in the lists developed by Russo et al.⁽⁶⁾ were intentionally left out.

After a previous selection of those words, the first version of the word bank contained 442 words, whose syllable templates were distributed as follows: 254 CV CV, 136 CVC CV, 36 CCV CV and 16 CV CCV.

2nd step: content validation

Content validation consisted of judgments of the 442 words from the word bank by expert and non-expert raters, in three distinct phases.

The expert raters were speech-language therapists who had a doctoral degree and worked as professors and researchers at universities in several regions of Brazil. They analyzed the words for familiarity and appropriateness.

Non-expert raters were professionals from different fields of knowledge and levels of education, who also analyzed the word bank. Other raters, who were also non-expert, played the role of listeners; they only listened to the words and performed an auditory recognition task through headphones in a soundproof booth⁽¹⁶⁾.

These expert and non-expert raters were chosen through convenience sampling. They received an e-mail message with information about the research project and an invitation to participate in it as raters. They confirmed their acceptance by submitting the document with their analysis of the words. The listeners were recruited through social media and local media and agreed to participate in the research by signing an Informed Consent Form (ICF).

In this step, data analysis was descriptive and used specific criteria for word choice, in each of the three phases of content validation, as described below.

Content validation: word familiarity

To begin the process of content validation, the 442 words were sent to 15 expert raters (seven speech-language therapists working in the field of Audiology and eight, in the field of Phonetics/Phonology) and ten non-expert raters who were supposed to rate each word for familiarity on a Likert scale style, as follows: extremely familiar (EF), very familiar (VF), familiar (F), hardly familiar (HF) or unfamiliar (UF).

In this step, the words rated by most raters as EF or EF + VF were selected.

Content validation: word appropriateness

Based on the results of the first rating (familiarity), a second list of words was developed and then sent to expert raters working in the field of Audiology. They were instructed to make a second rating, classifying each word as appropriate or inappropriate for the objective of the present study.

In order to classify each word as appropriate or inappropriate, the raters followed criteria such as phonetic aspects (place of articulation and voicing), ambiguous pronunciation and differences in familiarity across different socioeconomic levels and regions of Brazil, or emotional connotation.

Analysis of the data assessed by raters was based on frequency of occurrence of the classification adopted for each word. At the end of this stage, only the words considered as appropriate by most raters were maintained.

These words were used to create seven lists (LD-A, LD-B, LD-C, LD-D, LD-E, LD-F and LD-G) with 25 disyllabic words in each one of them. Although the words were not phonetically balanced, they were chosen for distribution into the different lists according to the following criteria: presence of a similar amount of the same phoneme in the different lists; even distribution of the disyllabic words with their different respective syllable structures; representation of phonemes of different frequency bands, based on the audiogram of familiar sounds of Brazilian Portuguese⁽¹⁷⁾.

The distribution of the phonemes by frequency bands of the audiogram considered the graph of the mean acoustic values of frequency and intensity of Brazilian Portuguese speech sounds, developed by Russo and Behlau⁽¹⁷⁾, thus guaranteeing the representation of low, medium, medium-high and high frequencies, which are necessary for adequate speech perception.

The word lists were recorded by a female speaker in a studio, according to standard ISO 8253-3:2012⁽¹⁸⁾, with a view to ensuring naturalness and uniform vocal quality and trying to avoid pronunciation with a marked regional accent while keeping the noise level below the test signal, at 40 dB.

The recording was made with a cardioid Neumann U87ai microphone, with pre-attenuation at to -10 dB, with a Pro Tools HD3 Accel recording system, *Digidesign* 192 digital recording interface running on a Mac Pro platform, AKG 55D headphones, Yamaha NS10M Studio Monitor. Sound editing

software: Sound Forge Pro 10; authoring software: Sony CD Architect 5.2.

The recorded material was equalized and digitally treated and edited with software, with a variation of ± 3 dB between all the items in the list. After that, it was recorded onto a Compact Disc (CD) and then played by means of a CD Player fitted to an audiometer.

Track 1 of the CD presented a reference signal of 1 kHz, and duration of 60 s, for adjustment of the unit volume (VU) meter of the audiometer. The second track contained a sentence for general guidance, with instructions on how the subjects were expected to perform the task (“You will hear a series of words and you have to repeat them the way you understand them. Repeat every word you hear”). The seven lists of disyllabic words were played on the other tracks. Each word was preceded by the introductory phrase: “Repeat the word”. There were regular time intervals between the words (4 s); such time was enough for the patient to respond and prepare for the next word⁽¹⁸⁾.

Content validation: auditory recognition of words

To conclude the content validation process, the seven lists of recorded words were presented at a level of 40 dBSL, for auditory recognition of the words, to 56 normal-hearing listeners with different educational backgrounds, with ages between 19 and 44 years, right-handed and speakers of Brazilian Portuguese.

The levels of education taken covered in this study were incomplete or complete elementary school, complete high school and complete higher education, as classified by Law No. 9,394 of December 20, 1996, which establishes the guidelines and bases of national education.

The evaluations were performed with 14 subjects at each level of education. They were instructed to listen and repeat the lists of recorded words, according to the instructions of the test. Half of them listened to the lists in the right ear and the other half in the left ear. The lists were presented in alternate order.

At this stage, data analysis covered the production of every single word. The words produced with more than one error were excluded and hence one list was deleted and the remaining words were rearranged into the other six lists (LD-A, LD-B, LD-C, LD-D, LD-E with 25 items each and combined in the same way as the seven previously arranged lists. They were edited on a new CD, based on the original recording, by a sound technician.

Step 3: equivalence search of the six lists of disyllabic words

The next step of development of the new speech material to determine SRPI was equivalence search of the six lists of disyllabic words which resulted from the content validation process. The aim of this step was to obtain evidence of reliability for the new test instrument.

To identify the most appropriate level of presentation of the six word lists for equivalence search, a pilot study was carried out with 12 subjects to achieve a percentage of 40% to 60% of correct production of disyllabic words, thus avoiding the floor effect (0%), or the “ceiling” effect (100%)⁽¹⁹⁾. Such result was

found when the words were presented at 26 dBHL, with ipsilateral speech noise, at 30 dBHL (signal-to-noise ratio of -4 dB).

Under this listening condition, the six lists of disyllabic words were presented to the subjects' ears by alternating the order of the ears and the sequence of presentation of the lists. In total, ten presentations were made in each position. We evaluated 60 normal-hearing subjects who were right-handed individuals, aged 19 to 24 years old and speakers of Brazilian Portuguese, recruited through local newspapers and social media, who agreed to participate in the study by signing an informed consent form. The subjects were instructed to listen and repeat the words.

The evaluations, both in the auditory recognition stage in the content validation process and in the equivalence search of the lists, used an Interacoustics AC 33 audiometer, TDH 39 earphones, and a Toshiba Compact Disc Player.

In this step, the data underwent descriptive and inferential analysis, based on the performance of the subjects to recognize the words in each list, in order to compare equivalence among the six lists of disyllabic words.

Inferential analysis was performed with the software Statistic 9.1, using the Wilcoxon test and the Friedman test (for multiple dependent samples, paired per subject), with a significance level of 5% (p -value ≤ 0.05).

RESULTS

In order to illustrate the steps of this research, Chart 1 shows a summary of the phases and results of the content validation process and the equivalence research, for development of the lists of disyllabic words.

In the 3rd phase of content validation, i.e., auditory recognition of words, there were 18 errors for 12 words. Among the incorrect productions, 8 words were produced incorrectly only once while 4, more than once, namely: /põtɕi/ as [fõtɕi], /sino/ as [seno], /vaka/ as [fakæ], /pizo/ as [pezo], /vila/ as [lilê], /xoŋko/ as [xombo], /xenda/ as [fendê], /venda/ as [fendê], /klube/ as [pluve] and [pluge], /nata/ as [nadê] three times, /kreme/ as [treme] three times, /krime/ as [kreme] twice.

Table 1 shows the results for auditory recognition, according to educational level.

Table 2 shows the results achieved in the stage of equivalence research, per list of disyllabic words.

Table 1 and Figure 1 shows the comparison among the six lists of disyllabic words in the equivalence search stage.

Chart 1. Results for content validation and equivalence search in the development of the lists of disyllabic words for speech audiometry testing

STEP	PROCEDURES	RESULTS
CONTENT VALIDATION (1st STEP) Assessment of word familiarity	442 disyllabic words 17 raters (9 experts – 5 audiologists and 4 phoneticians/phonologists; and 8 non-expert raters)	198 disyllabic words (44.8%) EF or EF+VF for most raters
CONTENT VALIDATION (2nd STEP) Assessment of word appropriateness	198 EF or EF+VF disyllabic words 5 expert raters (audiologists)	176 disyllabic words (88.89%) adequate for most raters
CONTENT VALIDATION (3rd STEP) Assessment of auditory recognition of words	176 disyllabic words Development and presentation of seven lists (LD-A, LD-B, LD-C, LD-D, LD-E, LD-F and LD-G) with 25 digitally-recorded disyllabic words to 56 listeners	172 disyllabic words (4 words excluded) 6 lists with 25 digitally-recorded (LD-A, LD-B, LD-C, LD-D, LD-E and LD-F) disyllabic words
EQUIVALENCE SEARCH Analysis of trustworthiness	Presentation of 6 lists of digitally-recorded (LD-A, LD-B, LD-C, LD-D, LD-E and LD-F) disyllabic words to 60 normal-hearing subjects	5 lists of equivalent digitally-recorded (LD-A, LD-B, LD-C, LD-D e LD-E), with 25 words each disyllabic words

Subtitle: EF = Extremely Familiar; VF = Very Familiar; LD-A = List of disyllabic words A; LD-B = List of disyllabic words B; LD-C = List of disyllabic words C; LD-D = List of disyllabic words D; LD-E = List of disyllabic words E; LD-F = List of disyllabic words F; LD-G = List of disyllabic words G

Table 1. Results for auditory recognition, in the content validation process, by educational level

Level of education (Law of Directives and Bases of National Education - 12/20/1996)	Subjects who made errors (n)	Incorrect productions (n)	Subjects who made more than one error (n)	p-value
ES (not completed)	5	8	3	0.459
ES	4	4	0	
HS	2	3	1	
HE	3	3	0	
TOTAL	14	18	4	

Statistically significant value ($p < 0.05$) - Kruskal-Wallis ANOVA Test

Subtitle: ES = Elementary School; HS = High School; HE = Higher Education

Table 2. Distribution of frequencies and descriptive measures of the results found in the stage of equivalence search, of the six lists of disyllabic words, after content validation

Lists		LD-A	LD-B	LD-C	LD-D	LD-E	LD-F
Percentage of correct productions	Less than 40%	2	0	5	5	4	6
	40-60%	57	53	54	51	55	53
	More than 60%	1	7	1	4	1	1
Total no. of subjects assessed		60	60	60	60	60	60
Descriptive measures	Minimum (%)	24	40	24	20	24	28
	Maximum (%)	64	68	64	64	64	64
	Standard deviation	7.89	7.94	7.89	8.8	8.1	7.4
	Mean	46.67	52.47	45.40	47.20	46.53	45.13
	Mode	40	48 and 60	40	40	40	40

Subtittle: LD-A = List of disyllabic words A; LD-B = List of disyllabic words B; LD-C= List of disyllabic words C; LD-D = List of disyllabic words D; LD-E = List of disyllabic words E; LD-F = List of disyllabic words F

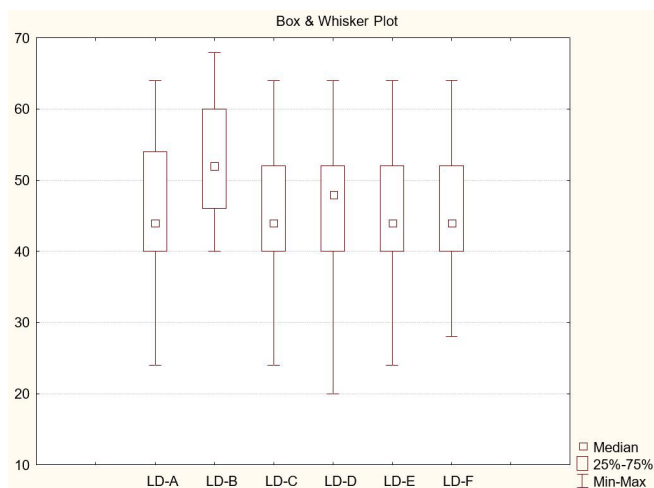


Figure 1. Representation of the comparison between the performance of the subjects in the auditory recognition of the LD-A, LD-B, LD-C, LD-D, LD-E and LD-F lists. Representation of the median and the minimum and maximum values in comparison to the percentage of correct productions of subjects per list of disyllabic words - Inter-list variability

Statistically significant values (p-value ≤ 0.05) Friedman’s test - multiple dependent samples, within-subject (paired); p-value = 0.00000

Subtittle: LD-A = List of disyllabic words A; LD-B = List of disyllabic words B; LD- C= List of disyllabic words C; LD-D = List of disyllabic words D; LD-E = List of disyllabic words E; LD-F = List of disyllabic words F; Median = Median; Min = Minimum; Max = Maximum

DISCUSSION

Because this work proposes a new test instrument to compose a set of speech materials available for assessment of SRPI, it is crucial to perform psychometric measurements and to determine their characteristics of validity and trustworthiness or reliability, so that this instrument can be used both in research and in clinical practice⁽²⁰⁾. Therefore, because there are few similar materials which have been used for such analyses, the present discussion will address the steps of the validation process performed in this research.

Validity analysis checks whether an instrument measures exactly what it is set out to measure. It can be performed through content validity, criterion validity and construct validity⁽²¹⁾. Content validity is described as a process of judgment, which consists in developing the instrument itself and evaluating it through expert analysis⁽¹⁶⁾.

Thus, since the beginning of the development of this speech material, there was a strict concern with the choice of words that would be submitted to the subsequent evaluation stages for the development of the proposed test instrument.

The choice of the items included common disyllabic words from the Portuguese lexicon, with paroxytone stress, because it is the most common stress pattern in Brazilian Portuguese⁽³⁾, and the syllable structures CV CV, CVC CV, CCV CV and CV CCV, which are the most frequent in the language⁽¹⁵⁾.

As suggested by experts⁽¹⁶⁾, the 442 Brazilian Portuguese words, selected according to previously established criteria, were carefully assessed by expert and non-expert raters for familiarity, appropriateness and auditory recognition (Chart 1). Thus, both qualitative and quantitative procedures were used⁽²²⁾. Particularly for materials developed for speech audiometry testing, uniform or homogeneous familiarity is required among the words or items of the test instrument^(4,23,24).

Thus, expert and non-expert raters made an analysis of familiarity and appropriateness of the words of the material. They maintained 176 words out of the 442 disyllabic items that had been initially selected, and these remaining words were submitted to auditory recognition (Chart 1).

In the incorrect productions found during auditory recognition, as described in the results, most errors involved changes to consonant sounds outnumbered those of vowels. This finding confirms the finding of a previous study⁽²⁵⁾, and corroborates the finding that speech intelligibility is more dependent on consonant sounds than on vowels⁽¹⁷⁾.

The incorrectly repeated words presented phonemes which are characteristic of bands of low, medium, high and medium-high range frequencies, in descending order, with approximate intensity values, ranging from 15 to 45 dB⁽¹⁷⁾.

However, the acoustic frequency and intensity of these words, including the words incorrectly repeated more than once, did not justify the errors in auditory recognition, because they

had more phonemes in the low and medium frequency bands, with intensity values between 25 and 45 dB⁽¹⁷⁾, hence they could have been recognized by the participating raters, all of whom were normal hearing subjects, who had air-conduction hearing thresholds up to 25 dBHL for frequencies between 0.25 and 8 kHz.

The 18 incorrect productions found in the stage of auditory recognition of words were made by 14 subjects of all levels of education, but mostly by subjects with incomplete or complete elementary school (ES). Four subjects made more than one error; three of them had incomplete ES level of education and one had completed higher education. However, there was no difference per level of education of the subjects participating in this study (Table 1).

This finding suggests that the choice of familiar words can minimize the effect of educational differences between the evaluated subjects and reinforces the need to include familiar words, because word familiarity is dependent on frequency of use in the language and is related to improved intelligibility of words⁽²⁴⁾.

The errors produced were therefore considered to be random, or influenced by the characteristics of the speaker's voice, or aspects relative to the recording and / or editing of the new test material under development, since these are some factors that interfere in auditory performance for recognition of the items that compose speech materials^(1,23,26).

Therefore, the word bank consisted only of EF or EF + VF words, which were deemed as appropriate in the opinion of most raters and were correctly recognized by listeners with different levels of education. This result ensured the requirements of familiarity and quite similar degree of difficulty, and mitigated the need of phonetic balance of words, because this element has been considered to be of secondary importance in the organization of speech materials^(6,12,24).

In order to collect evidence of reliability for the speech material proposed for SRPI, equivalence search was performed for the six lists made from the word bank with 172 disyllabic words which had resulted from the content validation process.

Trustworthiness, or reliability, is the accuracy and constancy of the measurements obtained when using a test instrument. It means that the instrument is faithful and obtains similar results in comparable situations⁽²⁰⁾.

Equivalence search of words used in speech materials for speech audiometry testing is carried out with several evaluation strategies. There are frequent descriptions of techniques with psychometric functions of intensity *versus* intelligibility of the proposed speech material^(13,27-29) and associated fixed noise⁽³⁰⁾.

There was no consensus in the literature as to the best research strategy to reach the objective in question. To evaluate the equivalence between the six lists of disyllabic words, a pilot study was carried out to define an evaluation situation that allowed the subjects to present speech recognition index other than 0% or 100%, avoiding the "floor effect" or the "ceiling effect"⁽¹⁹⁾.

As a result, the aim of the pilot study was to find percentage values between 40% and 60% of correct production of the disyllabic words, when the words were presented at 26 dBHL, with ipsilateral speech noise at 30 dBHL (4dB signal-to-noise ratio). When analyzing the equivalence of lists of Mandarin

monosyllabic words, other authors⁽²⁹⁾ also sought to obtain recognition scores between 40 and 60%.

In a first analysis, the performance of the subjects evaluated in the auditory recognition of words was found to be very similar for all the lists, except for the LD-B list. In the LD-A, LD-C, LD-D, LD-E and LD-F lists, most subjects had auditory performance between 40% and 60%, with a tendency to less than 40% (Table 2).

Descriptive measures have shown that these lists, in fact, are very balanced in terms of the intelligibility of each word. Similar minimum and maximum performance values, as well as mode values, coincide with the most common scores for these five lists. In addition, the mean and median, which represent a measure of central tendency for the speech recognition scores (Table 2), confirm the similarity between the LD-A, LD-C, LD-D, LD-E and LD-F lists (Figure 1).

This finding is in agreement with the assumptions about the criteria to be considered for development of speech materials, e.g., the same average difficulty and also an equal range of difficulty between the lists (26), unlike what occurred for list LD-B, in which there was a great number of subjects with a score above 60% and no subject with a score below 40%. This was evidence that the list was easier for auditory recognition than the other lists, and hence it was excluded.

Therefore, based on the assumption that the items of a speech material for audiometry testing cannot be either too easy or too difficult⁽¹³⁾, the five lists, in which the subjects presented a median performance in auditory recognition (between 40% and 60%) were maintained in the proposed test instrument.

These results, found through descriptive analysis, were confirmed by inferential analysis, thus reinforcing that only the LD-B list differed ($p = 0.00$) from the other lists (LD-A, LD-C, LD-D, LD-E and LD-F), as shown in Table 3 and in Figure 1.

There are a few reasons⁽¹⁾ why different lists of words can produce equivalent results, e.g., physical factors, such as the type of test stimuli; linguistic reasons, for example, word familiarity; phonetic factors, or aspects relative to recording and editing of

Table 3. Comparison between the six lists of disyllabic words in the equivalence search stage, based on subjects' performance, in the recognition of words per list

Lists	p-value
LD-A ≠ LD-B	0.00*
LD-A = LD-C; LD-D; LD-E; LD-F	> 0.05
LD-B ≠ LD-A; LD-C; LD-D; LD-E; LD-F	0.00*
LD-C ≠ LD-B	0.00*
LD-C = LD-A; LD-D; LD-E; LD-F	> 0.05
LD-D ≠ LD-B	0.00*
LD-D = LD-A; LD-C; LD-E; LD-F	> 0.05
LD-E ≠ LD-B	0.00*
LD-E = LD-A; LD-C; LD-D; LD-F	> 0.05
LD-F ≠ LD-B	0.00*
LD-F = LD-A; LD-C; LD-D; LD-E	> 0.05

*Statistically significant value ($p < 0.05$) - Wilcoxon Test

Subtitle: LD-A = List of disyllabic words A; LD-B = List of disyllabic words B; LD-C = List of disyllabic words C; LD-D = List of disyllabic words D; LD-E = List of disyllabic words E; LD-F = List of disyllabic words F

the test material and to the speaker's voice output, or even the presentation and execution of the test itself^{1,23,26}).

Thus, at the end of the equivalence research, it could be confirmed that the goal was reached in five of the six lists.

Importantly, the psychometric measures obtained are valid only for the recording of the lists used in this study. If the lists were recorded again by the same speaker or another speaker, or if there were reordered to produce different lists, new psychometric measurements would have to be taken.

It is noteworthy that after exclusion of a list (LD-B), the equivalent lists were renamed in alphabetical order and hence named LD-A, LD-B, LD-C, LD-D and LD-E.

In this work, it was found that the development and validation of the proposed digitally-recorded instrument reinforce previous statements from the literature about the design of speech materials and their respective forms of presentation. Thus, variability of result is reduced when determining SRPI.

However, monitored live voice is the predominant presentation form used in routine audiological practice in Brazil for speech audiometry testing; therefore, there may be some reluctance to use this evaluation instrument.

CONCLUSION

A word bank was developed with 172 disyllabic words, with satisfactory evidence of content validity, based on analysis of familiarity, appropriateness and auditory recognition of words. These words were distributed into six lists, with 25 disyllabic words each.

Five of the lists were considered to be equivalent, and showed satisfactory evidence of trustworthiness. The lists, called LD-A, LD-B, LD-C, LD-D and LD-E, were digitally recorded and made available on a Compact Disc for use in SRPI assessment, thus complementing the existing set of speech materials.

REFERENCES

1. Penrod J. Logoaudiometria. In: Katz J, editor. *Tratado de audiologia clínica*. 4. ed. São Paulo: Manole; 1999. p. 146-62.
2. Carhart R. Basic principles of speech audiometry. *Acta Otolaryngol*. 1951;40(1-2):62-71. <http://dx.doi.org/10.3109/00016485109138908>. PMID:14914512.
3. Chaves AD, Nepomuceno LA, Rossi AG, Mota HB, Pillon L. Reconhecimento de fala: uma descrição de resultados obtidos em função do número de sílabas dos estímulos. *Pro Fono*. 1999;11(1):53-8.
4. Carhart R. Problems in the measurement of speech discrimination. *Arch Otolaryngol*. 1965;82(3):253-60. <http://dx.doi.org/10.1001/archotol.1965.00760010255007>. PMID:14327024.
5. Menegotto IH, Costa MJ. Avaliação da percepção de fala na avaliação audiológica convencional. In: Boéchat EM, Menezes PL, Couto CM, Frizzo ACF, Scharlach RC, Anastasio ART, editores. *Tratado de audiologia*. Rio de Janeiro: Guanabara Koogan; 2015. p. 67-75.
6. Russo ICP, Lopes LQ, Brunetto-Borginanni LM, Brasil LA. Logoaudiometria. In: Santos TMM, Russo ICP, editores. *A prática da audiologia clínica*. 6. ed. São Paulo: Cortez; 2007. p. 135-54.
7. Eldert E, Davis H. The articulation function of patients with conductive deafness. *Laryngoscope*. 1951;41(9):891-909. PMID:14874534.
8. Mendel LL, Mustain WD, Magro J. Normative data for the Maryland CNC test. *J Am Acad Audiol*. 2014;25(8):775-81. <http://dx.doi.org/10.3766/jaaa.25.8.7>. PMID:25380123.
9. Gates GA, Feeney MP, Higdon RJ. Word recognition and the articulation Index in older listeners with probable age-related auditory neuropathy. *J Am Acad Audiol*. 2003;14(10):574-81. <http://dx.doi.org/10.3766/jaaa.14.10.6>. PMID:14748554.
10. Mendel LL, Owen SR. A study of recorded versus live voice word recognition. *Int J Audiol*. 2011;50(10):688-93. <http://dx.doi.org/10.3109/14992027.2011.588964>. PMID:21812631.
11. Sá G. Análise Fonética da Língua Portuguesa falada no Brasil e sua aplicação à Logoaudiometria. *Rev Bras Med*. 1952;9(7):482-90. PMID:13064475.
12. Mangabeira-Albernaz PL. Logoaudiometria. In: Pereira LD, Schochat E, editores. *Processamento auditivo central: manual de avaliação*. São Paulo: Lovise; 1997. p. 37-42.
13. Harris RW, Goffi MVS, Pedalini MEB, Merrill A, Gygi MA. Reconhecimento de palavras dissilábicas psicometricamente equivalentes no português brasileiro faladas por indivíduos do sexo masculino e do sexo feminino. *Pro Fono*. 2001;13(2):249-62.
14. Canongia MB. Manual de terapia da palavra, anatomia, fisiologia, semiologia e o estudo da articulação e dos fonemas. 3. ed. Rio de Janeiro: Livraria Atheneu; 1981. 543 p.
15. Viaro ME, Guimarães-Filho ZO. Análise quantitativa da frequência dos fonemas e estruturas silábicas portuguesas. *Estudos Linguísticos*. 2007;36(1):27-36.
16. Pasquali L. *Psicometria: teoria dos testes na psicologia e na educação*. 4. ed. Rio de Janeiro: Vozes; 2011. 399 p.
17. Russo ICP, Behlau M. *Percepção da fala: análise acústica do português brasileiro*. São Paulo: Lovise; 1993. 57 p.
18. ISO: International Organization For Standardization. ISO 8253-3:2012: acoustics: audiometric test methods: part 3: speech audiometry. Geneva: ISO; 2012.
19. Thornton AR, Raffin MJM. Speech-discrimination scores modeled as a binomial variable. *J Speech Hear Res*. 1978;21(3):507-18. <http://dx.doi.org/10.1044/jshr.2103.507>. PMID:713519.
20. Marques-Vieira CMA, Sousa LMM, Carvalho MLR, Veludo F, José HMG. Fidelidade e validade na construção e adequação de instrumentos de medida. *Enformação*. 2015;5:25-32.
21. Alexandre NMC, Coluci MZO. Validade de conteúdo nos processos de construção e adaptação de instrumentos de medidas. *Cien Saude Colet*. 2011;16(7):3061-8. <http://dx.doi.org/10.1590/S1413-81232011000800006>.
22. Hyrkäs K, Appelqvist-Schmidlechner K, Oksa L. Validating an instrument for clinical supervision using an expert panel. *Int J Nurs Stud*. 2003;40(6):619-25. [http://dx.doi.org/10.1016/S0020-7489\(03\)00036-1](http://dx.doi.org/10.1016/S0020-7489(03)00036-1). PMID:12834927.
23. Hood JD, Poole JP. Influence of the speaker and other factors affecting speech intelligibility. *Audiology*. 1980;19(5):434-55. <http://dx.doi.org/10.3109/00206098009070077>. PMID:7436861.
24. Owens E. Intelligibility of words varying in familiarity. *J Speech Hear Res*. 1961;4(2):113-29. <http://dx.doi.org/10.1044/jshr.0402.113>. PMID:13731816.
25. Longone E, Borges ACC. Principais trocas articulatórias envolvidas na obtenção do índice percentual de reconhecimento de fala em indivíduos portadores de perda auditiva neurosensorial. *Acta AWHO*. 1998;17(4):186-92.
26. Egan JP. Articulation testing methods. *Laryngoscope*. 1948;58(9):955-91. <http://dx.doi.org/10.1288/00005537-194809000-00002>. PMID:18887435.

27. Nissen SL, Harris RW, Jennings LJ, Eggett DL, Buck H. Psychometrically equivalent Mandarin bisyllabic speech discrimination materials spoken by male and female talkers. *Int J Audiol.* 2005;44(7):379-90. <http://dx.doi.org/10.1080/14992020500147615>. PMID:16136788.
28. Wang S, Mannell R, Newall P, Zhang H, Han D. Development and evaluation of Mandarin disyllabic materials for speech audiometry in China. *Int J Audiol.* 2007;46(12):719-31. <http://dx.doi.org/10.1080/14992020701558511>. PMID:18049961.
29. Han D, Wang S, Zhang H, Chen J, Jiang W, Mannell R, Newall P, Zhang L. Development of Mandarin monosyllabic speech test materials in China. *Int J Audiol.* 2009;48(5):300-11. <http://dx.doi.org/10.1080/14992020802607456>. PMID:19842805.
30. Ji F, Xi X, Chen AT, Zhao WL, Zhang X, Ni YF, Yang SM, Wang Q. Development of a mandarin monosyllable test material with homogenous items (II): Lists equivalence evaluation. *Acta Otolaryngol.* 2011;131(10):1051-60. <http://dx.doi.org/10.3109/00016489.2011.583267>. PMID:21599549.