Testes de rastreamento x testes de diagnóstico: atualidades no contexto da atuação fonoaudiológica***

Screening versus diagnostic tests: an update in the speech, language and hearing pathology practice

Bárbara Niegia Garcia de Goulart*
Brasília Maria Chiari**

* Fonoaudióloga. Doutora em Ciências pela Universidade Federal de São Paulo - Escola Paulista de Medicina. Professora Titular da Feevale. Endereço para correspondência: Rua São Manoel, 2107/302 - Porto Alegre - RS - CEP 90620-110 (bgoulart@via-rs.net).

**Fonoaudióloga. Professora Titular do Departamento de Fonoaudiologia da Universidade Federal de São Paulo - Escola Paulista de Medicina.

***Trabalho Realizado no Departamento de Fonoaudiologia da Universidade Federal de São Paulo.

Artigo de Atualização

Artigo Submetido a Avaliação por Pares

Conflito de Interesse: não

Recebido em 10.03.2006.
Revisado em 03.06.06; 15.03.2007; 25.05.2007.
Aceito para Publicação em 25.05.2007.

Abstract
Background: evaluation instruments, properties, selection indicators, application and validation of screening and diagnostic tests. Aim: to present some concepts concerning screening and diagnostic tests and their application according to a specific purpose. To present a few practical examples of the application of these instruments related to human communication, as well as to present validation criteria of tests in the population and criteria used for the rational selection of screening or diagnostic instruments in health programs and health services based on epidemiological concepts indexed in Scielo, Lilacs or Medline up to January 2007. Conclusion: diagnostic instruments differ from screening instruments in their objectives and eligibility criteria. Sensibility and specificity are two important indicators to be considered when choosing an instrument for screening or diagnosis. Reproducibility, time required to complete the evaluation and previous preparation of patients, if needed, are also indicators to be considered when choosing an instrument. Publication and information exchange regarding the properties of evaluation instruments, used for diagnosis or screening, related to the Speech, Language and Hearing Sciences must be systematically stimulated. Besides that, improving the knowledge about methodologies and evaluation instruments under different perspectives contribute to the better use of human and financial resources. Furthermore, the elaboration of studies that promote the correct validation of screening and diagnostic instruments used in human communication disorders contributes to the increase in knowledge in the field of Speech, Language and Hearing Sciences and, indirectly, to the acknowledgement of this science, based on technical-scientific evidence, in health promotion.
Key Words: Diagnostic; Screening; Sensitivity; Specifity; Evaluation.

Resumo
Tema: instrumentos diagnósticos, propriedades e uso de indicadores para seleção, aplicação e validação de instrumentos de diagnóstico e de rastreamento. Objetivo: apresentar conceitos ligados aos instrumentos de avaliação e sua aplicação de acordo com o objetivo, seja rastreamento ou diagnóstico. Também são apresentados alguns exemplos práticos de aplicação de instrumentos de avaliação ligados à comunicação humana, bem como critérios de validação de testes na população e critérios utilizados para sua escolha e aplicação racional em serviços e programas de saúde a partir de pressupostos epidemiológicos pesquisados em artigos indexados nas bases de dados Scielo, Lilacs ou Medline até janeiro de 2007. Conclusão: instrumentos de avaliação e diagnóstico clínico diferem dos instrumentos de rastreamento em relação a seus objetivos e critérios de elegibilidade. São indicativos da precisão de um instrumento de avaliação, seja para rastreamento ou diagnóstico, a sensibilidade e a especificidade de tal instrumento. Questões como reprodutibilidade, tempo para realização do teste ou exame e preparação prévia do paciente também devem ser considerados quando da seleção de instrumentos de avaliação clínica. O conhecimento e disseminação de informações ligadas às propriedades dos instrumentos de avaliação ligados a Fonoaudiologia devem ser incentivados sistematicamente. Além disso, a ampliação da gama de conhecimentos acerca das diferentes perspectivas ligadas às metodologias e instrumentos diagnósticos contribuem com a melhor racionalização de recursos humanos e financeiros. A elaboração de estudos que promovam a validação dos instrumentos correntemente utilizados para rastreamento e diagnóstico dos distúrbios da comunicação humana colabora com o avanço do conhecimento ligado a Fonoaudiologia e, indiretamente, para com o reconhecimento da ciência fonoaudiológica, baseada em evidências técnico-científicas, na promoção da saúde.
Palavras-Chave: Diagnóstico; Rastreamento; Sensibilidade; Especificidade; Avaliação.

Referenciar este material como:
GOULART, B. N. G.; CHIARI, B. M. Screening versus diagnostic tests: an update in the speech, language and hearing pathology practice (original title: Testes de rastreamento x testes de diagnóstico: atualidades no contexto da atuação fonoaudiológica). *Pró-Fono Revista de Atualização Científica*, Barueri (SP), v. 19, n. 2, p. 223-232, abr.-jun. 2007.

Introduction

In the last decades, allied to the immense technological advancement, biosciences has been developed a lot of knowledge, mainly in the last few years. Along with this process, the role of Speech and Hearing Sciences has been recognized and research related to the human communication has been developed progressively, contributing with a big part of the knowledge that is used nowadays. In the present time, studies that show stronger scientific evidences, that bring more evidences of speech-language and hearing pathology (SLP) impact in the prevention and rehabilitation of human communication earns more space in the SLP research agenda, in order to be possible earn and maintain SLP work in the health related teams, contributing to the integrality of health promotion.

In this sense, is extremely important to guarantee and extend the arguments, knowledge and practices that are already diffused in other areas of knowledge in order to contribute with the advancement of the knowledge related to the evaluation, diagnosis and handling of human communication and its related disorders.

It is necessary to think that health workers commonly use some kind of measure at the diagnostic and / or at the therapeutic procedures, like the examples referred by Noble et al (2004) and Gomes (2005). Therefore, when researching the clinical history of the patient or the history of his complaint we measure the time elapsed since the beginning of the signs and symptoms, its variation and intensity in this period, between others. Beyond that, the maximum phonation time measurement, facial thirds measures, accompaniment of each swallowing stage, measurement of the auditory sharpness and speech disorders evaluation are some practical examples of measurement use of tests and instruments in the SLP clinical approach.

Based in some conceptions, these data contribute to the SLP diagnosis and therapeutic planning, considering that treatment planning shall be based on more accurate diagnosis and it all can improve the success and effectiveness of SLP intervention.

Another example, described by Oliveira et al (2002), remits us to the deafness diagnosis. The authors affirm that nowadays is consensual that just medical observation and the family suspicion are not sufficient for the detection of deafness in the first year of life. More truthful evaluation is desirable, based on standardized instruments such as the otoacustic emissions and BERA (Brain Electrical Response Audiometry), between others.

Stands out that more precise knowledge of evaluation instruments properties, as screening or diagnostic tests, exams or clinical evaluations a link should be made in order to apply all new technological resources must be used rationally. This is just as well that more and more these technologies are available in quantity , but their parameters are not always properly well known as it should be to decide what type of instrument shall be used (Krauss-Hisses, 2004; Potvin, 2005).

Decisions about asking for complementary evaluations and the use of other exams that vary from low to high complexity should be harnessed, between others, to the characteristics of each test and/or procedure and their contribution for diagnosis, or even, bring more subsidies for the therapeutic planning.

As an example we can use a hypothetical speech disorder as a complaint in a 6 years old child. It is well known that such evaluation can be made by spontaneous speech data collection and recording using pictures that represent all of the Brazilian Portuguese phonemes or even from the emission of the words by the SLP for subsequent repetition by the child in evaluation (Castro and Wertzner, 2006; Goulart and Ferreira, 2006).

In the present time, we have a lot of instruments in Brazil for screening and/or diagnosis of most of communication disorders. Although, it is necessary that all of these instruments be validated between subjects with and without the disorders that intend to diagnose in order to evaluate with more precision the potentials of the instrument that is utilized. The translation of the characteristics of the tests and/or exams of diagnosis and/or screening can be characterized by the measurement of the sensitivity, specificity and predictive values

of each approach, test or exam.

This article presents some concepts, reflections and application of screening tests and diagnostic instruments in the scope of human communication healthiness and communication disorders. Beyond that, some pertinent points related to screening or diagnostic tests and instruments validation, based on Epidemiologic concepts will be discussed.

These points will be discussed based in the scientific literature published until February 2007 in Scielo, Lilacs and/or Medline. Bibliographic research used the MESH or DECS terms: rastreamento, track, screening, screening tests, screening test, screening procedures, screening method, test diagnosis, diagnosis, standardized test, tests validation, evaluation, sensitivity, specificity, validity, results reliability, evaluation of biomedical technology, promotion and prevention in health, health promotion and language evaluation.

Characteristics and properties of evaluation Instruments

Costa et al (2004), Jeckel et al (2005a) and Hochman et al (2005) refer two distinct objectives that stand out in the assessment based on an evaluation instrument in clinical activities or for research: precision and accuracy.

Accuracy refers to the capacity of a measure be correct in on average. When the measure is not accurate, it is biased. The precision, also known as reliability, is the capacity of a measure to give a very similar result (or the same result) in repeated measurements of a same fact. Both qualities are very important and should be assured. If one of these is absent, data (or results) become useless, therefore it could be not trustworthy.

Accessing the facts regarding to reliability, to the accuracy, the importance rank of the test for decision-making, test costs, risks and the acceptability of the instrument and its contribution in the healthiness reestablishment or in the adverse effects caused by the instrument also should be considered on the decision making to use these instruments (Newman et al, 2003; Noble et al, 2004; Potvin, 2005). Such concepts are close-related to evidence-based clinical practice. These concepts were introduced in the clinical practices in North America since the 70's, mainly in the Medicine and nowadays these arguments have been highlighted in many discussions related to speech-language and hearing sciences action in Brazil (Goulart, 2002b; Goulart e Chiari, 2006).

In this way, when considering clinical studies or population based surveys findings to evidence-based decision making in order to health promotion trough collective actions or in the clinical setting all methods and other situation related to the research read shall be considered. Methods used to select sample participants, instruments used and their background, training or abilities needed for those who applied the instruments or protocols used in the study (Potvin, 2005; Goulart, 2002b). Also, it is imperative to remember that only data based on similar populations can be overstepped and, when there is an extremely selected sample, with a lot of exclusion criteria, these subjects may not represent the population in general. Those factors can interfere in the results related to tests, instruments or approaches for diagnosis or screening that show satisfactory results in specially designed research settings. So, they may not present similar answers in other opportunities.

How important an instrument can be to contribute for the decision-making regarding the therapeutic planning, including itself the relation of cost-benefits of the instrument application, the risks to which the professional and/or the patient will be submitted by the use of such instrument and the acceptability of the procedure by the patient (and his relatives, when in case) also should be considered when deciding about the use of these instruments. These criteria are widely described and discussed by the scientific community since the 70's and more recently join in these administrative arguments of the health related sectors, users of health services and the professionals that often use such instruments in their routine (McKinlay and Marceau, 2002; Zuckerman, 2003; Georg et al, 2005).

Krauss-Silva (2004) refers about technologies and procedures coverage policies, as well as the elaboration of clinical

guidelines and health services evaluation and management are increasingly dependent on formal technology assessments in developed countries. Regarding Brazil, this study identified some operational barriers to the elaboration and utilization of technology assessment that are related to the adequacy and availability of relevant data as well as to the absence of enough trained researchers and decision-makers.

Krauss-Silva (2004) refer to the health technology assessment (HTA) concept in for assign those approaches used to select among the existing alternatives, individual technologies or technological joints for deal with a similar health related need and/or problem. Although more limited, HTA are also helpful to select technological assemblies organized as programs (a program can match up to several assemblies of technologies) to manage different health related problems, either in the technical phase of priorities selection and for planning and management of health services as a evidence-based critic subsidy to organize clinical guidelines.

Characteristics and application of screening and diagnosis instruments

Screening aim is the identification of an illness or risk factor not already recognized neither by the patient clinical history, physical exam, laboratory test or other procedure that can be applied quickly (Bhopal, 2002; Newman et al, 2003; Toscano, 2004). Screening instruments separate people that are apparently well, but have an illness or a risk factor for a disease of those that do not present them.

Patients screening in the illness pre-clinical phase is already common in the medical practice and, actually, screening tests to detect some diseases are being extended for the population in general, without complaints (Georg et al, 2005; Fletcher & Fletcher, 2006).

However, it is important be conspicuous that the present recommendations related to the prevention of illness lead to the sense of submit the population to specific tests according to the predominance of characteristic disturbances related to age groups, sex and clinical characteristics (Bhopal, 2002; Newman et al, 2003; Fletcher & Fletcher, 2006).

Coutinho et al (2003), Toscano (2004) and Georg et al (2005) affirm that the application of screening tests should be done considering the premise that earlier diagnosis can contribute to better prognostic and earlier treatment, improving the chances to avoid comorbities and irreversible symptoms. These considerations have been already made in some studies published by Fiorindo et al (2004), Toscano (2004) and Cardoso et al (2004), related to many specialties related to health sciences.

Toscano (2004) and Fletcher & Fletcher (2006) refer that a screening test should detect most number of cases with the lowest cost possible, beyond do not cause adverse reactions or side effects. The authors also consider that there is less benefit between screening tests if they are compared to diagnostic tests to plan the treatment and solve most of the symptoms. However, screening tests must be more secure and have less adverse reactions than the current employed clinical tests. Its use is translated in low precision regarding the diagnosis, considering that diagnostic instruments are more often specific.

An ideal screening instrument should last just a few minutes to be carried out, beyond require a minimum of patient's prior prepare and do not depend on special pre-schedule, beyond cause no or minimum discomfort to the patient or his family (Toscano, 2004; Fletcher & Fletcher, 2006). Whatever possible it is recommended to use a cheap screening test, easy to apply and that causes none or minimum discomforts and/or trouble to the patients (Toscano, 2004; Jeckel et al, 2005b). All of these authors also refer that the results of a screening test should be valid, reliable and useful.

Briefly, the validity of an screening or diagnosis instrument can be measured by its ability to do what it propose, categorize adequately the individuals with "pre-clinical" symptoms of a disease as positive test and those subjects without pre-clinical symptoms of the illness like negative test. These relations between illness and positive test, not-illness and positive test, illness and negative test and not-illness and negative test are express through sensibility, specificity and

predictive values of the tests, even for screening or diagnostic tests (Fletcher & Fletcher, 2006).

The accuracy of the instrument depends, not only just of the sensibility or specificity but, also, it depends on the prevalence of the illness in the evaluated population using these instruments. Rarest the illness is, more specific should be the test to be helpful, to detect the cases (illness people). However, if an illness has a high prevalence in the population tested (taking itself as an example the diabetes mellitus and the auditory hearing loss that started after the seventh decade of life), the test should have high sensibility to be helpful in the clinical practice, therefore, otherwise, a negative result will express a false result. A good screening test should have high sensibility to not loose the cases of illness in the population tested, as well as it must have high specificity in order to reduce the number of people with false-positive results that need subsequent investigation (Bhopal, 2002; Fletcher & Fletcher, 2006).

Diagnostic instruments are commonly more accurate and bring more details related to the disease that they are able to diagnose. As much as the screening instruments, the diagnostic tests need to be valid to the populations to which are destined, including healthy people with the same characteristics that the group with a disorder or disease that it is designed for. Beyond that, on the occasion of the decision-making by the utilization of an instrument for diagnosis should be considered its cost-benefit relation, its capacity to add information that contribute with the therapeutic planning and complement, if necessary, the diagnosis carried out from the anamnesis and clinical exam finds.

For the diagnostic instruments actually contribute with information to others already given data it is important that they have high specificity, so, rarely the test will be positive in the absence of illness.

Krauss-Silva (2004) refers that the international studies frequently focus populations whose cultural, demographic and genetic characteristics differ of the majority of the Brazilian population, what is extremely important when evaluating the straight application of these health technologies in the Brazilian reality. The author detaches that such differences can modify, in a significant way, important parameters related to the accuracy, the efficacy and the utility or preference of such instruments according to the different realities that we have in all the regions of Brazil. Validity problems of the instruments measures can influence either the adhesion to these technologies, as well its effectiveness in our community.

Characteristics and validation of screening or diagnostic instruments

For that a screening or clinical diagnostic instrument be current used and generate useful and trustful data it must be submitted to a validation procedure.

Instruments are validated according to the population to which they are designed to (adults, infants, with or without deficiencies), they should be previously tested to verify the sensibility and specificity in healthy people and in those already sick for which such instruments are designed to, even to diagnostic or screening procedures.

Beyond that, studies for measure instruments validation should count on re-evaluation by an instrument for the same purpose properly already validated or, in case of impossibility to use an instrument with these characteristics, the gold standard should be a clinical evaluation with an expertise.

To estimate the sensibility of an instrument of measure it is necessary to count on a number of persons that present the event in which we are interested (disorder, alteration, illness) and introduce them in this group. From this, it is necessary to verify in how many of these the instrument could detect the event in study (positive test), as describes Gomes (2005).

In this case, the sensibility of the instrument is measured by the proportion of cases in that the results of the test in study, when positive, agree with the results of a test considered for reference (gold standard) and the specificity is the proportion of cases in the test, when negative, be in agreement with the test considered for reference (Gomes, 2005; Hochman et al, 2005; Fletcher & Fletcher, 2006).

Taking the use of a screening instrument to detect speech disorders in infants from 7

years and older as an example, called TERDAF (Teste de Rastreamento de Distúrbios Articulatórios de Fala), designed and validated by Goulart (2002a), to which were submitted 200 infants with and without such disorder. If the test detects speech disorder in 163 infants, the sensibility of the test will be 81.5%, in other words, this is the possibility of TERDAF to be positive in an infant with a speech disorder.

The sensibility of a diagnostic or screening test (or exam) measures the capacity of this to detect something that exists (Gomes, 2005; Hochman et al, 2005). For example, a sensitive deafness diagnostic instrument (considering both together, equipment and examiner) is the one that, with high probability, detects deafness in a deaf patient.

In this way, an instrument is considered more sensible when increase the probability of detection of an event (outcome, disorder or illness).

The specificity is characterized by the capacity of an instrument to not detect nonexistent events (Gomes, 2005; Hochman et al., 2005). In case of people with normal arterial tension, an instrument with high specificity would detect that the arterial tension in this group is in the limits of normality. Taking itself an example linked to the human communication, a specific instrument would not detect voice disorders in a person without dysphonia.

Regarding to the specificity, it is necessary to submit a certain number of individuals that do not present the characteristic of interest to the instrument in study to verify how many will present negative result for the event related to the instrument in study.

Maintaining the previous example, if we submitted 220 children without speech disorders to TERDAF screening test and among those the referred instrument presented negative result in 99 of them, to specificity of the instrument would be 45%. It means that for each ten children without speech disorders submitted to the screening test, approximately four would present negative result (negative test).

Another thing to be considered concerns to the cut point of the test, it is related to the level (or degree) of alteration that, when detected, turns the test result as positive. This choice involves a decision between increase the sensibility of the instrument at the expense of specificity reduction, or vice versa. Nearly all of the researchers should carefully evaluate the relative importance of the instrument sensibility and specificity to establish the most adequate cut point to define the diagnostic transition.

As general strategy, when the main worry is to avoid the false-positive result (for example, the test result can induce to the indication of a risky procedure to the patient), then the cut point should turn to have the more specificity and if the biggest worry is to stay away from false-negative result (the test result of a patient with suspicion of dysphagia, for example), the cut point should consider the higher sensibility (Fletcher & Fletcher, 2006). The authors say that the ROC (Receive Operator Caracteristic) Curve is the best way to establish the cut point, optimizing the sensibility and specificity of the tests.

To build the ROC Curve the researcher should select several points or levels of test alteration (different intensity levels, for example) and establish the sensibility and the specificity in each point. Following, a sensibility graphic is made, based on the proportion of the false-positive results of the test. The ideal test is the one that achieves the upper left extremity of the graphic. One of the advantages of this approach is that the curves (or graphics) of different tests can be compared.

Newman et al. (2003) agree that much better the test, more nearby will be its curve of the left upper corner of the graphic. These authors also remember that the necessary elements for a diagnostic or screening study to be considered consistent are subject to answer three main questions: a) the existence of independent and blind comparison between the test in study and the "gold standard test" that is current used to diagnosis the illness. Authors describe that the patients sample should be submitted to the test in study and the gold standard (of reference test), used as supply of that the patient has or does not have the disease that the tests intend to diagnose. Beyond that the results of a procedure cannot be already known by the person who applies or interprets the test that is being compared, seen that those two elements are essential to avoid biased interpretations, conscious or

unconscious, of the investigators. The diagnostic test (in study) should be applied in an appropriate sample of similar patients to that found in the heath services. The patients in study should present clinical findings commonly presented by ones with the disease in study, as well clinical characteristics related to the differential diagnosis of the disease. And, finally, the authors suggest that the diagnostic test should be validated in a second independent group of patients. The first study (with the diagnostic test) can predict the accuracy of the test in study. To confirm the test performance it is necessary a new search, with an independent sample, when the accuracy can be measured and checked again.

Whenever is possible, the reference test should be enough known and, ideally, the best one that already exist. Usually this instrument is called the 'gold standard'. It can be an exam of the same kind of the one that is being studied or even another type of instrument. Using as an example, we can compare the sensibility and specificity of the otoacustic emissions (OAE) with the BERA (Brain Electrical Response Audiometry) to detect deafness in newborns.

Didactically, we can summarize the comparison between an instrument in study and a gold standard test in a two-by-two table (or contingencies table), presented in the table 1.

The 2x2 Table is widely used in epidemiological studies and its possibilities of use related to human communication research examples were described by Goulart (2002b).

Therefore, (a + c) is the total of subjects with positive gold standard, it enclose all the people that we already know that have the characteristic (or the disease) in study. The sum (b + d) correspond to the group of people that do not present the characteristic in study.

The cell (a) of the table hold all the people that have positive results even in the test in study and in the gold standard, these are also called the "true positive" group. In the cell (b) are allocated those who do not present the event in study according to the gold standard and present positive test when submitted the evaluation by the instrument in study. The group in the cell (b) is named "false positive".

In the same way, we call "false negative" people arranged in the cell (c) and "true negative" the group of patients arranged in the cell (d).

As described by Gomes (2005), sensitivity is calculated by the reason between the number of true positive and the total of positive gold standard test and specificity is characterized by the reason between the all those patients with true negative test and all with negative gold standard test (see the schema below):

Sensitivity = a / (a + c) Specificity = d / (b + d)

Reliability of diagnostic and/or of screening tests (instruments) also should be considered. This test property is related to the possibility that the test could be applied by any professional (or not, according with its objective) and patients in different moments allowing to have similar test results. A reliability decrease can be resultant of problems with the own instrument, such as technical application difficulties that occur when complex procedures demand a special expertise by the examiner or different interpretations of the test results. As an example, we could refer to worse BERA results when analyzed by different examiners or videonasopharyngoscopy changes pre and post speech-language therapy as an indicator of velopharyngeal closure improvement.

A paper about evaluation instruments validation approach should bring the reliability that can be expected of the instrument. This description is especially important when analyzing the instrument results, mainly, when the instruments demand some experience of the examiner, as in videolayngoscopy, videonasopharyngoscopy or videofluoroscopic analysis. Thus, it is possible to confirm the agreement level between the findings (or analyses) when there are divergences between the examiners responsible by the interpretation of such results. Reliability level inter and intra-examiner should be analyzed either, such as described in the Tinnitus Handicap Inventory (THI) for the Brazilian Portuguese validation, carried by Ferreira et al (2005).

When the reliability of an instrument in study seems to be good, but inter-examiner disagreement occurs, it is necessary to

consider if the couple, examiner and instrument, are good in order to discriminates effectively those with and without the disease in study. Nevertheless, this instrument can be helpful in clinical evaluation, therefore, if the test sensitivity is higher than the inter-examiner agreement, and there is a high prevalence of disagreement between inter-examiners, the instrument persist to be simple and reliable, but dependent of those who interpret its results (Jaeschke et al, 1995). The authors also describe that if two type of instruments offer similar answers, the application of the second test will offer very short or none contribution to the diagnosis process. This matter should be considered when constant need of resources rationalization is required, especially in the contemporary Brazilian society, as previously discussed (Goulart, 2003; Noble et al, 2004).

Oliveira el al. (2002) discuss that deafness screening tests for infants at risk is also controversial, mainly regarding to the instruments and methods that can be used. In this way, while some authors defend the achievement of otoacustic emissions (OAE) followed by BERA, others advocate the utilization of BERA as a first choice test. The authors still refer that in any of the situations it is necessary to pay attention to the continuous evolution of techniques and knowledge, therefore new plans and new approaches, as the Products of Distortion and the automatic BERA, shortly can assume positions of highlight and, either, first choice instruments.

Another example close to speech-language and audiology pathologist action regard to the evaluation instruments choice is described by Hage et al. (2004). The authors mention in this paper instruments with different characteristics and objectives to child development evaluation. They detach that the development scales, which reflect the main steps during infantile development and have the aim to determine preschooler specific development level. From these, the evolution level of the infant is obtained through facts related about his/her development and from the straight observation of his/her behavior. Some scales assess specifically the language development, like the Early Language Milestone Scale (ELMS) and Reynell Developmental Language Scale (RDLS), other evaluate several aspects of the development, among these, Hage et al. (2004) detach Denver Developmental Screening Test (DDST) and Escala de Desenvolvimento Comportamental de Gesell e Amatruda (EDCGA).

Moreira e Ferreira-Junior (2004) carried out a comparative study between people with noise-induced hearing loss (NIHL) and individuals with normal hearing and concluded that the Percentage Index of Speech Recognition (PISR) was not helpful in the diagnosis of NIHL.

Study published by Marteleto e Pedromônico (2005) established the validity of criteria of the Inventory of Autistic Behaviors (IAB) and the authors found that in the sample submitted to the test, 81.6% of the infants with autism were correctly identified. The authors also describe that the Inventory of Autistic Behaviors (IAB) showed low sensitivity (57.89%) and high specificity (94.73%) when the cut point became 68. When the cut point was diminished to 49 points, the scale sensitivity increased to 92.1% and the specificity continued on the average (92.6%). In this way, the authors suggest that IAB is a promising instrument for identify autism in infants, especially with a cut point of 49, also in the clinical and educational contexts.

A study published by Ferreira et al. in 2005 describe in details the validation of the Tinnitus Handicap Inventory (THI) in the Brazilian population, beyond the application and results found in the validation of the instrument in the population researched.

TABLE1. Data disposition in a contingencies table.

	Gold Standard +	Gold Standard –	TOTAL
Test in Study +	a	b	a + b
Test in Study –	c	d	c + d
TOTAL	a + c	b + d	N = a + b + c + d

Conclusion

Deeply knowledge about speech-language and hearing needs in the different population strata (age group, sex and others, according to specific disorders related to these variables) is vital to improve the effectiveness of efforts, financial and human resources and come, whenever possible to develop collective health. Therefore, as it was described by Krauss-Silva (2003) the chronic incipiencies of activities and application of the health technologies evaluation have been visible enough and expressed in the financial policy and procedures coverage, either by the public health or in the roll of procedures of the supplementary health (health insurance). Then, for that the speech-language and audiology therapy extends its insertions in the health related services it is necessary that more and more all of the professionals have knowledge and be able to use the available tools for evidence-based actions.

But, for that that occur, research centers and universities should still contribute strongly to the validation and enlargement of knowledge about the application of the available technologies related to the human communication.

Beyond that, knowledge production should be added , as much as possible, to broad disclosure in scientific communication vehicles, meetings and journals also to speech-language pathologists, audiologists and undergrad students through the enclosure of these subjects in the pedagogical projects of under and post graduation programs, as already occurs in many scientific areas.

The importance of take into account the tested population sample for instruments validation concerning the human communication should consider the regionalisms regarding the speech and language evaluation; age group and occupation or professional activity regarding voice or orofacial motricity assessment, between others. In this way, comprehensive description of the examined population and the criteria employed for the diagnosis in an instrument validation research are fundamental to the professional that will use the instrument or test in the clinical setting or in another research. Satisfactory instruments validation and knowledge of its technical properties in the Brazilian population also can add more subsidies to the priorities selection and delineation of the health related programs content by the decision-makers in different levels of management.

On the same way, when most of the epidemiological details of the population that need speech-language and audiology care are known, and their different ways of care needs based on regional realities, efficacy, cost-effectiveness, cost-benefit, as well systematic reviews would be necessary. And to carry on these type of studies, it would be necessary to combine many studies, but with similar methodology. Such studies can give more consistent evidences for the advancement of speech-language and audiology action, from the progress of the science regarding to the human communication.

Although, it is necessary to consider that neither all of the health related disorders, even connected to the human communication are subject to prevention and that only the opinion of speech-language and audiology services users do not constitute sufficient, neither satisfactory evidences based on the scientific and methodological point of view for the decision-making about the technologies and resources that would be employed in the speech-language and audiology care.

Garber (2001) and Krauss-Silva (2004) mention some studies about the irrational use of technologies and procedures based on the users' point of view. These studies show some evidences about technological use without evidence of benefit and vice versa, technology use in adverse conditions, when efficacy or accuracy can be not the same and different application of technologies without the corresponding variation to have better results. They also describe the big fascination exercised by new technologies, not always better and nearly always more expensive, beyond commonly complementary or similar to the ones that already exist.

It is important that the speech-language therapist and audiologist is aware of these questions and, whenever necessary, search for subsidies that can contribute with strong evidences to decisions taken about the use of tools and instruments in their work.

Reference

BHOPAL, R. S. Natural history, spectrum, iceberg, population patterns and screening: interrelated concepts in the epidemiology of disease. In: _____. *Concepts of Epidemiology.* Oxford, Oxford University Press, 2002. 317 p. Disponível em: <http://www.pitt.edu/~super1/lecture/lec18491/001.htm>. Acesso em: 05 maio 2006.

CARDOSO, A. A.; MAGALHÃES, L. C.; AMORIM, R. H. C.; PAIXÃO, M. L.; MANCINI, M. L.; ROSSI, L. D. F. Validade preditiva do movement assessment of infants para crianças pré-termo brasileiras. *Arq. Neuro-Psiquiatr.,* São Paulo, v. 62, n. 4, p. 1052-1057, dez. 2004.

CASTRO, M. M.; WERTZNER, H. F. Estimulabilidade e tipos de erro de fala. *R. Soc. Bras. Fonoaudiol.*, São Paulo, v. 11, n. 1, p. 1-9, jan.-mar. 2006.

COSTA, R. M. S.; VELASCO, T. R.; COSTA, L. S.; RODRIGUES, M. L. V. Eficácia da associação de métodos diagnósticos no rastreamento do glaucoma. *Arq. Bras. Oftalmol.,* São Paulo, v. 67, n. 1, p. 37-41, jan.-fev. 2004.

COUTINHO, E. S. F.; HUF, G.; BLOCH, K. V. Ensaios clínicos pragmáticos: uma opção na construção de evidências em saúde. *Cad. Saúde Públ.*, Rio de Janeiro, v. 19, n. 4, p. 1189-1193, ago. 2003.

FERREIRA, P. É. A.; CUNHA, F.; ONISHI, E. T.; BRANCO-BARREIRO, F. C. A.; GANANÇA, F. F. Tinnitus Handicap Inventory: adaptação cultural para o Português Brasileiro. *Pró-Fono R. Atual. Cient.,* Barueri (SP), v. 17, n. 3, p. 303-310, set.-dez. 2005.

FLETCHER, R. H.; FLETCHER, S. W. Prevenção. In: FLETCHER, R.; FLETCHER, S. *Epidemiologia clínica:* elementos essenciais. 4. ed. Porto Alegre: Artes Médicas, 2006. p. 179-201.

GARBER, A. M. Evidence-based coverage policy. *Health Aff.,* Bethesda (USA), v. 20, n. 5, p. 62-82, sep.-oct. 2001.

GEORG, A. E.; DUNCAN, B. B.; TOSCANO, C. M.; SCHMIDT, M. I.; MENGUE, S.; DUARTE, C.; POLANCZYK, C. A.; GRUPO DE TRABALHO DE AVALIAÇÃO DA CAMPANHA NACIONAL DE DETECÇÃO DE DIABETES MELLITUS. Análise econômica de programa para rastreamento do diabetes mellitus no Brasil. *R. Saúde Públ.,* São Paulo, v. 39, n. 3, p. 452-460, jun. 2005.

GOMES, U. A. Características dos instrumentos de medida. In: FRANCO, L. J.; PASSOS, A. D. C. (Org). *Fundamentos de Epidemiologia.* Barueri (SP): Manole, 2005. p. 151-164.

GOULART, B. N. G.; FERREIRA, J. Articulatory Speech Disorders Screening (TERDAF) - validation in to 1st grade public school students population. *Ci. Saúde Col.*, Rio de Janeiro, supl. esp., p. 5636-5636, 2006.

GOULART, B. N. G. A fonoaudiologia e suas inserções no Sistema Único de Saúde: análise prospectiva. *R. Fonoaud. Bras.,* Brasília, v. 2, n. 4, p. 29-34, dez. 2003.

GOULART, B. N. G. *Teste de rastreamento de distúrbios articulatórios de fala em crianças da 1a série do ensino fundamental público.* 2002a. 70 f. Dissertação (Mestrado em Epidemiologia) - Departamento de Medicina Social, Universidade Federal do Rio Grande do Sul, Porto Alegre.

GOULART, B. N. G. Contribuições da epidemiologia para a pesquisa e atuação clínica em fonoaudiologia. *R. Fono Atual,* São Paulo, v. 21, n. 3, p. 60-63, jul.-set. 2002b.

HAGE, S. R. V.; JOAQUIM, R. S. S.; CARVALHO, K. G.; PADOVANI, R.; GUERREIRO, M. M. Diagnóstico de crianças com alterações específicas de linguagem por meio de escala de desenvolvimento. *Arq. Neuro-Psiquiatr.,* v. 62, n. 3a, p. 649-653, set. 2004.

HOCHMAN, B.; NAHAS, F. X.; OLIVEIRA FILHO, R. S.; FERREIRA, L. M. Desenhos de Pesquisa. *Acta Cir. Bras.*, São Paulo, v. 20, supl. 2, p. 2-9, 2005.

JAESCHKE, R.; GUYATT, G. H.; SACKETT, D. L.; THE EVIDENCE BASED MEDICINE WORKING GROUP. How to Use an Article About a Diagnostic Test. Based on the Users' Guides to Evidence-based Medicine and reproduced with permission from JAMA. *Jama,* v. 271, n. 5, p. 389-391, 1994 (artigo previamente publicado) and v. 271, n. 9, p. 703-707, 1994. Copyright 1995, American Medical Association. Disponível em < http://www.cche.net/text/usersguides/diagnosis.asp>. Acesso em: 25 fev. 2007.

JECKEL, J. F.; ELMORE, J. G.; KATZ, D. L. Delineamentos comuns de pesquisa usados em epidemiologia. In:_____. *Epidemiologia, bioestatística e medicina preventiva.* 2. ed. Porto Alegre: Artmed, 2005a. p. 80-112.

JECKEL, J.; ELMORE, J.; KATZ, D. Métodos de prevenção secundária. In: _____. *Epidemiologia, bioestatística e medicina preventiva.* 2. ed. Porto Alegre: Artmed, 2005b. p. 236-247.

KRAUSS-SILVA, L. Avaliação tecnológica e análise custo-efetividade em saúde: a incorporação de tecnologias e a produção de diretrizes clínicas para o SUS. *Ci. Saúde Col.*, Rio de Janeiro, v. 8, n. 2, p. 501-520, 2003.

KRAUSS-SILVA, L. Avaliação tecnológica em saúde: questões metodológicas e operacionais. *Cad. Saúde Públ.*, Rio de Janeiro, v. 20, supl. 2, p. S199-S207, 2004.

McKINLAY, J. B.; MARCEAU, L. D. The end of the golden age of doctoring. *Int. J. Health Serv.,* Amityville (USA), v. 32, n. 2, p. 379-416, 2002.

MARTELETO, M. R. F.; PEDROMÔNICO, M. R. M. Validade do inventário de comportamentos autísticos (ICA): estudo preliminar. *R. Bras. Psiquiatr.,* São Paulo, v. 27, n. 4, p. 295-301, dez. 2005.

MOREIRA, R. R.; FERREIRA-JUNIOR, M. Testes de fala: aplicação em portadores de perda auditiva induzida por ruído. *Pró-Fono R. Atual. Cient.,* Barueri (SP), v. 16, n. 3, p. 293-300, set.-dez. 2004.

NEWMAN, T. B.; BROWNER, W. S.; CUMMINGS, S. R. Delineando estudos de testes médicos. In: HULLEY, S. B. et al. *Delineando a pesquisa clínica:* uma abordagem epidemiológica. 2. ed. Porto Alegre: Artmed, 2003. p. 203-224.

NOBRE, M. R. C.; BERNARDO, W. M.; JATENE, F. B. A prática clínica baseada em evidências: parte III - avaliação crítica das informações de pesquisas clínicas. *R. Assoc. Med. Bras.*, São Paulo, v. 50, n. 2, p. 221-228, jan.-abr. 2004.

OLIVEIRA, P.; CASTRO, F.; RIBEIRO, A. Surdez infantil. *R. Bras. Otorrinolaringol.,* São Paulo, v. 68, n. 3, p. 417-423, maio 2002.

POTVIN, L. Why we should be worried about evidence-based practice in health promotion. *R. Bras. Saúde Mater. Infant.,* Recife, v. 5, supl. 1, p. S93-S97, dez. 2005.

TOSCANO, C. M. As campanhas nacionais para detecção das doenças crônicas não-transmissíveis: diabetes e hipertensão arterial. *Ci. Saúde Col.*, Rio de Janeiro, v. 9, n. 4, p. 885-895, out.-dez. 2004.

ZUCKERMAN, D. Hype in health reporting: "checkbook science" buys distortion of medical news. *Int. J. Health Serv.*, Amityville (USA), v. 33, n. 2, p. 383-389, 2003.