

Research strategies and evidence-based practice in relation to communication disorders

Dalia M. Osman

Phoniatric Unit, Otolaryngology Department, Faculty of Medicine, Cairo University, Cairo, Egypt

Correspondence to Dalia M. Osman, Assistant Professor of Phoniatrics, Otolaryngology Department, Faculty of Medicine, Cairo University, Cairo, Egypt
e-mail: dmostafa9999@yahoo.com

Received 25 May 2012

Accepted 25 June 2012

The Egyptian Journal of Otolaryngology
2012, 28:162–170

Introduction

To be effective, all professional practices, including speech and language pathology, should be built on a foundation of basic and applied research. Clinicians working in these fields, similar to clinicians in other fields, need to be aware of the validity and reliability of the methods they select while dealing with their patients. Moreover, the current emphasis on evidence-based practice in medicine across all scientific fields emphasizes the importance of understanding the philosophy of science, research strategies, and evidence-based practice. This would aid in bridging the gap between the research and clinical work, thereby, maximizing the therapeutic benefit that the patients gain.

Objectives

The aim of this article is to review some of the facts related to research methodology, research evaluation, and evidence-based practice as well as to highlight the importance of integrating individual clinical expertise with the best available external evidence obtained from systematic research.

Recommendations

(a) Research should be considered as an integral part of clinical practice. (b) Keeping the concept of validity, reliability, and evidence-based practice in mind at all times while designing scientific studies. (c) Discussing the project design with an advisor or a group of colleagues to help ensure that validity is maintained at every stage of the process. (d) Correctly forming a research problem, sticking to research ethics while carrying out research work, and closely reviewing the selected literature before as well as during the study. (e) Randomly selecting a representative sample population of adequate size to represent the target population. (f) Identifying variables before starting research, choosing appropriate study designs, and appropriately dealing with missing data and confounding (extraneous) variables.

Keywords:

evidence-based practice, research and communication disorders, research methodology, research strategies, study designs

Egypt J Otolaryngol 28:162–170
© 2012 The Egyptian Oto - Rhino - Laryngological Society
1012-5574

Objectives

The aim of this article is to review some of the facts related to research methodology, research evaluation, statistical designs, and evidence-based practice as well as to highlight the importance of integrating individual clinical expertise with the best available external evidence obtained from systematic research.

Review of literature

Science, theories, and hypotheses

Science is a step-by-step acquisition of knowledge. The goals of science are to describe natural events or phenomena, understand and explain natural phenomena, and control natural phenomena by understanding the causes of events and predicting their occurrences [1].

Speech and language science is based on many foundations, among which are theories and hypotheses. A theory

is a comprehensive description and explanation of a total phenomenon. However, a hypothesis involves a more specific prediction stemming from a theory. As such, hypotheses are limited in scope compared with theories. For example, the behavioral theory of language learning explains the process of learning in all children around the world. A hypothesis might address language learning in children with autism. To test their hypothesis, scientists gather data that are obtained by systematic observation (empirical, based on events that resulted in some form of sensory contact) and in many cases, experimentation. Scientists observe events and record some measured values of those events (e.g. the actual number of dysfluencies when stress is increased) [2].

Research and its types

Research can be defined as the structured inquiry that utilizes acceptable scientific methodology to solve problems and create new knowledge that is generally acceptable. Research is what scientists do as they practice

science. It is the process of asking and answering questions. It is science in action [3].

Research can be classified in different ways: from application perspectives, objectives perspectives, and modes of enquiry perspective. From the application perspective, research can be classified into pure and applied research. From the objectives perspective, it can be classified into descriptive, exploratory, correlation, and explanatory research. From perspective of the mode of enquiry, research can be classified into quantitative and qualitative research [4]. The most commonly used types of research in the field of communication disorders are experimental and descriptive researches [3].

Experimental research

The hallmark of experimental research is the investigation of cause-effect relationships [1]; for example, studying the efficacy of a language intervention program on the child’s academic achievement. This lends itself to a pretest/post-test methodology in which the research will determine the academic achievement before the intervention and then again after the language intervention program has been implemented. However, to determine the actual impact of an intervention, a pretest/post-test methodology must always be compared with a control group [5].

Experimental studies must entail random assignment of units; for example, people to the levels or categories of the manipulated variable [6,7]. The goal of having these two groups is to show that the experimental participants improved and the control participants did not, thus showing the efficacy of treatment. In forming two or more groups, researchers use either randomization or matching. Using the first option, they randomly draw a sample, or a small number, of participants required for the study from population. A population is a large, defined group (e.g. patients scheduled for laryngectomy surgery, individuals who stutter) identified for the purpose of a study. Randomly, selected participants are then randomly assigned to different groups. These two kinds of randomization – random selection and assignment – are expected to result in groups that are equal to begin with. The selection is random when each potential participant in the population has an equal chance of being selected for the study. The assignment is random when two levels of randomization reduce experimenter bias in selecting participants and ensures that the sample is representative of the population.

Quasi experiments refer to investigations that have all the elements of an experiment, except that the participants are not randomly assigned into groups [8].

After assigning participants to groups, experimenters need to manipulate independent variables to assess the effect of these variables on independent variables [1]. A good experimental research also involves conditions that are carefully controlled to eliminate extraneous or confounding variables to ensure that only the independent variable of interest is affecting the dependent variable [3]. A confounding variable is an extraneous variable that is

statistically related to (or correlated with) the independent variable. This means that as the independent variable changes, the confounding variable changes along with it. Failing to take a confounding variable into account can lead to a false conclusion that the dependent variables are in a causal relationship with the independent variable [9] (Table 1).

Descriptive research

The main objective of a descriptive research is to describe a certain phenomenon. However, it cannot establish a cause-effect relationship as there is no manipulation of variables. Descriptive research studies include comparative research, normative research, correlation research, and ethnographic research.

The purpose of comparative research is to measure the similarities and differences in groups of individuals with defined characteristics. Here, the confounding variables are not controlled [10], for example, patients with dementia might perform differently on receptive tasks than healthy individuals because of educational or socio-economic differences rather than the presence or absence of dementia. In addition, the variables are not termed independent and dependent variables. They are referred to as classification variables (e.g. having or not having a history of dementia) and criterion variables (receptive scores).

Correlation research is another type of descriptive research. It measures the strength of the relationship or associations between variables. However, it does not imply causation [11]. A positive correlation means that as one variable increases, the other also increases. However, the negative correlation means that as one variable increases, the other decreases [12]. An example of correlation descriptive studies would be studying the correlation between autistic features and sensory integration dysfunction or communication skills in a group of children with autism.

Developmental (normative) research is a descriptive research that measures changes in participants over time as individuals become older. It can be longitudinal, cross-sectional, or semilongitudinal. Cross-sectional studies are simple in design and are aimed at determining the prevalence of a phenomenon, problem, attitude, or issue by

Table 1 Examples of independent and dependent variables in speech and language pathology researches [3]

Independent variables	Dependent variables
Therapy	Improving communication skills
Stressful situation	Amount of stuttering
Amount of noise in environment	Amount of difficulty hearing speech
Length and complexity of reading passage	The number of errors a client with aphasia makes while reading it
Number of times client uses hard glottal attack	Amount of dysphonia present when the patient is speaking
Frequency of the use of hearing aids	Hearing outcome
Treatment	Effect of treatment

taking a snap shot or a cross-section of the population. This obtains an overall picture as it stands at the time of the study [13]. In cross-sectional studies, participants from various age levels are selected and studied [3]. An example of a cross-sectional study would be taking a sample from third graders, fourth graders, and fifth graders and comparing their language to see how language develops with age, assuming that when the third graders grow, they will have the same language skills as fourth graders and thus, because it is based on an assumption, it is less accurate than the longitudinal study but saves much time, effort, and cost.

In longitudinal (cohort) studies, the same participants are studied over time [14]. Some longitudinal studies last several months, whereas others can last decades. In longitudinal studies, variables are not manipulated and no causal relationships are detected [13]. Although lengthy and expensive, longitudinal studies are more accurate than cross-sectional studies in describing a naturally occurring phenomenon [3]; for example, stages of pragmatic development in typically developing children.

A semilongitudinal study is a sort of a compromise between cross-sectional and longitudinal studies. In semilongitudinal studies, the total age span to be studied is divided into several overlapping age spans. The participants selected are those who are at the lower end of each age span, and they are followed until they reach the upper end of their age span. For example, groups might be as follows: 3-year-olds followed until the age of 4 and 4-year-olds followed until the age of 5, and 5-year-olds followed until the age of 6. The researcher can then make observations both between as well as within participants as time passes [3].

Cohort (longitudinal) studies can be further subdivided into retrospective and prospective studies [14]. Retrospective/*ex post facto* means after-the-fact research [3]; that is, it examines information and specimens that have been collected in the past [14] (e.g. an attempt to determine how many children admitted to a children's hospital in the past 5 years had a swallowing disorder).

In contrast to retrospective research, prospective studies begin in the present and follow participants in the future [14]; for example, one might design a study involving children who attend the outpatient clinic for postcochlear implant rehabilitation.

Ethnographic research is a descriptive research. It is relatively new in the field of communication disorders. It involves observation and description of a naturally occurring phenomenon by dealing with qualitative data. The disadvantages of ethnographic research are that it is time consuming, often expensive, yields data that are difficult to quantify, and lacks the objectivity of experimental research [3]. An example of ethnographic research would be studying how the production of the affricate /dʒ/ varies in children from upper Egypt versus ones living in Cairo.

A survey research examines the prevalence of a certain phenomenon by questioning individuals. The tools most

commonly used are questionnaires and interviews. These need to be designed carefully to avoid any possible bias [3]. The purpose of survey research is to generate a detailed inspection of the prevalence of phenomena in an environment by asking individuals as opposed to direct observation [10].

Research process

Research consists of three steps: posing a question, collecting data to answer the question, and presenting an answer to the question [15]. These steps can be further divided into many substeps, among which are as follows.

Choosing a topic

For a researcher to choose a topic, it is important to consider a broad area of inquiry and interest. This may be as broad as 'language', but it should be an area that is of interest to the researcher. However, a broad area is useful only at the beginning of a research plan.

Within a broader topic of inquiry, each researcher must begin narrowing the field into a few subtopics that are of greater specificity and details. Oftentimes, students as well as professional researchers discover their topics in a variety of conventional and unconventional ways. Many researchers find that their personal interests and experiences help to narrow their topic. The researcher also has to consider whether it would be feasible to collect the data, and if so, would it be ethical, valid, and reliable to conduct [5]. In the field of communication disorders, for example, a researcher might be interested in 'language' but could focus more specifically on 'language development in children'. Although this topic is still too broad for a research project, it is more focused and can be further specified into a coherent project.

Formulating a research problem

A good research question has to address an important relevant issue. It has to be logical, ethical, feasible to study, and novel. This means that there will be some new aspect of the study that has never been examined before. This does not mean that we should avoid replicating past research. In fact, not only is replication a good way to obtain a research methodology, it is how science is supposed to advance knowledge. However, when replicating a previous study, it is best to add or change one or two things to increase the novelty of the research.

A good research question needs to be 'operationalizable': oftentimes, beginning researchers pose questions that cannot be operationalized or assessed methodologically using research instruments. In general, the more abstract the idea, the harder it is to operationalize.

The research question also has to be of adequate cost-effectiveness value. It also needs to be within a reasonable scope: the more focused the research question, the more likely it will be a successful project [5]. For example, a study that seeks to identify the prevalence of autism in a specific area is more likely to succeed than a comparable study that seeks to identify the prevalence of autism in the world population.

Planning a research

In designing a study, the researcher may find it helpful to consider the relationship between the research question (the question he wants to answer), the study design, and what is the study expected to answer, taken into consideration the anticipated errors of implementation. Good judgment by the investigator and advice from colleagues are required for the many trade-offs involved and for determining the overall viability of the research. Estimation of the sample size is also one of the most important early parts of planning a research [14].

Conducting a research

Selecting a population: Once the researcher has chosen a hypothesis to test in a study, the next step is to select the study population from the target population. A target population refers to all participants of interest to whom the conclusions of the study will be applied. However, study population refers to the individuals actually available and accessible for study [12].

A researcher often cannot work with the entire population of interest, but instead, must study a smaller sample of that population in order to draw conclusions about the larger group from which the sample is drawn [16]. An example of a population is the population of children with stuttering in Egypt, an example of a sample is a group of third-grade Egyptian children with stuttering as opposed to those in the second grade, and an example of an element is a single child with stuttering. In selecting the study population, researchers may need to specify research entry data using inclusion criteria, exclusion criteria, and stratification [14].

Collecting data: Data comprise observation on one or more variables; any quantity that varies is termed a variable (e.g. variables affecting a study on fundamental frequency levels can be age, sex, etc.). Data are usually obtained from a sample of individuals that represents the population of interest [11].

The data collected from the study population can be quantitative or qualitative [12]. Qualitative (categorical or nominal) data are verbal descriptions of attributes of events. In nominal data, a category is present (e.g. hypernasality) or absent (normal nasality) [3].

Numerical/quantitative data are numerical descriptions of attributes of events. Examples of quantitative data are when the researcher states that in a 5-min spontaneous speech sample, the participants omitted word final phonemes 75% of the time. However, data can be described as discrete when the variable can only take certain whole numerical values; for example, days on stuttering [11].

An ordinal scale is a numerical scale that can be arranged according to rank orders or levels. Ordinal scales use relative concepts such as greater than or less than. The intervals between numbers of categories are unknown. Examples of ordinal scales of measurement are 1 = strongly agree, 2 = agree, 3 = neutral, 4 = disagree, and 5 = strongly disagree [3].

An interval scale of measurement is a numerical scale that can be arranged according to rank orders or levels; the numbers on the scale must be assigned in such a way that intervals between them are equal with respect to the attribute being scaled. The ratio scale has the same properties as the interval scale, but numerical values must be related to an absolute zero point. The zero suggests an absence of the property being measured [17]. An example of a ratio scale is one that involves frequency counts in stuttering; it is possible to have zero instances of stuttering in a speech sample.

Sometimes, an arbitrary value such as a score is used when quantities cannot be measured. For example, a series of questions in a sensory integration questionnaire are summed up to obtain the overall tactile/vestibular sensory dysfunction score.

Data entry: While entering data, it is essential to avoid errors and missing data. For categorical data, assigning numerical codes to categorical data before entering the data will be essential. The researcher has to thoroughly revise the data to exclude the existence of any outliers. There are several reasons why participants could be outliers. One reason is that they are different from other participants and another reason is that participants have systematically responded without really thinking of what they were doing.

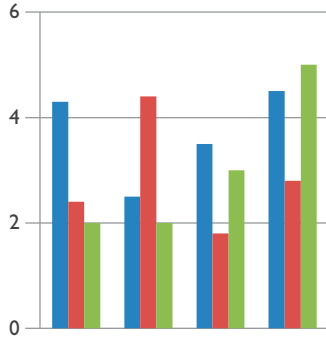
Cleaning of data is a rather simple but necessary step. The researcher needs to check that all data lie within the expected range; for example, calculating the mean scores for each item and then checking that the listed values lie within the expected range [8].

Diagrammatic representation of data: Diagrams can help researchers get a 'feel' for the data. Diagrams are often powerful tools for conveying information about the data, for providing simple summary pictures, and for spotting outliers and trends before any formal analyses are performed. Charts can be in the form of pie charts, bar graphs (bar charts), segmented column, histogram, line graphs, scatter plots, stocks, or doughnuts (Figs 1–8).

Carrying out statistical analysis: Statistics encompass the methods of collecting, summarizing, analyzing, and drawing conclusions from the data. The aim of any statistical study is to condense data in a meaningful way and extract useful information from them. It is true that a part of the theory of statistics involves effective ways of summarizing and communicating masses of information that describe some situation. This part of the overall theory and set of methods is usually known as descriptive statistics.

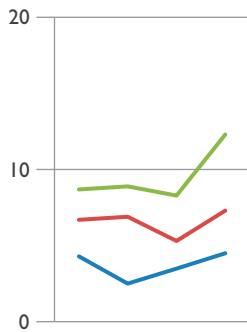
Descriptive statistics: Descriptive statistics simply describe the data pertaining to a population or a sample, specifically the center of the data (e.g. mean, median, and mode), spread (variability of the data points), and shape of the plotted graph (e.g. symmetrical or not). There must be some evidence that the sample chosen is a representative of the target population as this will considerably affect the interpretation of the results obtained [11].

Figure 1



Diagrammatic representation of the data using column graph.

Figure 2



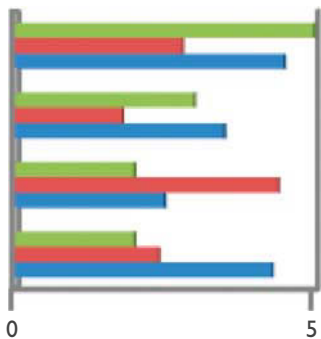
Diagrammatic representation of data using line graph.

Figure 3



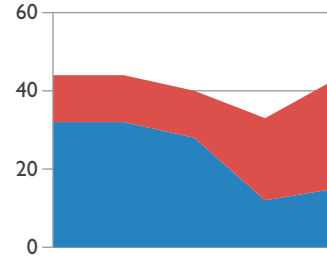
Diagrammatic representation of data using pie chart graph.

Figure 4



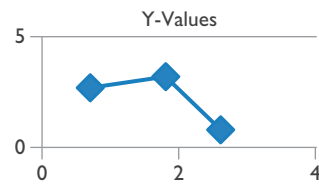
Diagrammatic representation of data using bar graph.

Figure 5



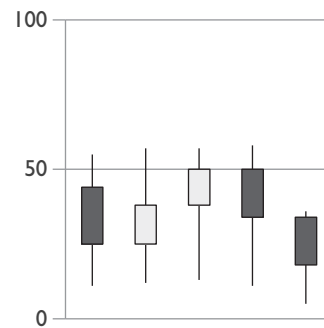
Diagrammatic representation of data using area graph.

Figure 6



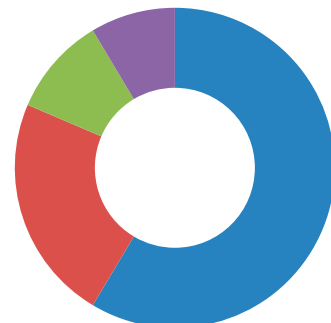
Diagrammatic representation of data using X-Y scatter graph.

Figure 7



Diagrammatic representation of data using stock graph.

Figure 8



Diagrammatic representation of data using doughnut graph.

Describing the average values of data:

- (1) The arithmetic mean: Often called the mean of a set of values is calculated by adding up all the values and dividing this sum by the number of values in the set.
- (2) The median: If data are arranged in order of magnitude, starting with the smallest value and ending with the largest value, then the median is the middle value of this ordered set.
- (3) The mode: The mode is the value that occurs most frequently in a data set; if the data are continuous, the data are usually grouped and the modal group is calculated [11].
- (4) The weighted mean (overall mean): When a weighted mean is used, certain values of the variable of interest are more important than the others.
- (5) Trimmed mean: It is the one in which the highest and the lowest values are omitted, thus reducing the distorting effect or outliers. A 5% trimmed mean is one in which the top 5% and the lowest 5% of the data are removed.
- (6) Approximate mean: It resembles the weighted mean, but is used when data points are intervals.
- (7) Geometric mean: It summarizes changes over time as the average ratio or rate of change [8]; for example, in tracking how fast the practice of speech and language pathology grown in Egypt in the last 3 years.

Describing the spread of data/measures of dispersion:

If there are two summary measures of a continuous variable, one that provides an indication of the average and one that describes the spread of the observation, then the data can be condensed in a meaningful way.

- (1) Range: High and low scores.
- (2) The SD: It is the distance from the mean. The SD is the square root of the variance. We can consider the SD as a sort of average of the deviations of the observations from the mean.
- (3) Ranges derived from percentile: The values of x that divide the ordered set into 10 equally sized groups, that is, 10th, 20th, etc., are called deciles. The values of x that divide the ordered set into four equally sized groups, that is, 25th, 50th, and 75th percentiles, are called quartiles. The 50th percentile is the median. Using percentiles, we can obtain a measure of spread that is not influenced by outliers by excluding the extreme values in the data set and determining the range of the remaining observations.
- (4) The variance: One way of measuring the spread of data is to determine the extent to which each observation deviates from the arithmetic mean. The larger the deviation, the greater the variability of the observation.
- (5) If the distribution of data is relatively symmetrical, the three measures of central tendency should be the same. This would result in a normal distribution or a bell-shaped curve [3].

Inferential statistics: Although descriptive statistics form an important basis for dealing with data, a major part of

the theory of statistics is concerned with how can one go beyond a given set of data and make general statements about the large body of potential observations, of which data collected represent but a sample. This is the theory of inferential statistics [18].

Writing a research report: The researcher then writes his research report with its entire components; for example, abstract, introduction, objectives, results, and discussion. Research reports need to be comprehensive as well as readable. This needs to be followed by drawing conclusions, deriving possible implications, and recommending future studies. While writing research reports, researchers need to adequately refer to recent references whenever possible.

Evaluation of a research

Measures in speech–language pathology, whether they apply to research studies or clinical practice, need to be valid and reliable [3]. Validity refers to whether or not a study is well designed and provides results that are appropriate to generalize to the population of interest [5]. It is an indicator of how much meaning can be placed upon a set of test results. For example, a valid child language test should measure language skills, not auditory memory [3].

Any research should be evaluated on the basis of two distinct features: internal validity and external validity. Internal validity supports the conclusion that the causal variable caused the effect variable in a specific study [19]. Internal validity applies in studies that seek to establish a causal relationship between two variables, and it refers to the degree to which a study can make good inferences about this causal relationship. The essence of internal validity is whether or not a researcher can definitively state that the effects observed in the study were in fact because of the manipulation of the independent variable and not because of another factor.

‘Third variables’ that the researcher may not consider or may not be able to control can affect the outcome of a study and can therefore prevent internal validity [5]; for example, determining the effect of a drug intake on voice and proving that the drug was the cause of voice change and not (e.g. using non-sex-matched groups). Thus, the researcher cannot prove whether the differences between the two groups were secondary to differences in the male–female distribution across the two groups or because of a true effect of the independent variable (the drug) on the dependent variable (voice change).

Good experimental techniques in which the effect of an independent variable on a dependent variable is studied under highly controlled conditions (including eliminating of any confounding variable) usually allow for higher degrees of internal validity than, for example, single case designs [10]. Unfortunately, many factors that can reduce the internal validity include instrumentation, history, statistical regression, maturation, attrition, testing, participant selection biases, and interaction of factors [3].

External validity refers to generalizability; that is, to what settings, populations, treatment variables, and

measurement variables the effect can be generalized [20] (i.e. it is concerned with the extent to which the conclusions can be generalized to the broader population). A study is considered to be externally valid if the researcher's conclusions can in fact be accurately generalized to the population on a large scale [16] (i.e. across time and space). External validity is usually divided into two distinct types, population validity and ecological validity (whether the results can be applied in real-life situations), and they are both essential elements in determining the strength of an experimental design [19].

The external validity of a study may be affected by several factors. These include the Hawthorne effect (the extent to which the participants' knowledge that they are taking part in a research or that they are being treated differently than usual), participant selection, multiple treatment interference, and reactive and interactive effects of pretesting; for example, individuals who regularly abuse their voices by speaking loudly and using hard glottal attacks might fill out a questionnaire before treatment that assessed the frequency with which they used such abusive vocal habits. The participants, thus sensitized to how often they abused their voices, might begin to modify their vocal quality [3].

Face validity is a measure of how representative a research project is 'at face value', and whether it appears to be a good project. In contrast, construct validity defines how well a test or an experiment measures up to its claims [19]. A test designed to measure speech nasality must only measure that particular construct, not closely related ideals such as voice quality. Construct validity is the degree to which test scores are consistent with theoretical constructs or concepts. For instance, a test of language development in children should meet the theoretical expectation that as children grow older, their language skills improve [3].

Convergent validity tests whether the constructs that are expected to be related are, in fact, related, whereas discriminant validity tests that constructs that should have no relationship do, in fact, not have any relationship (also referred to as divergent validity) [19].

Other types of validity that are important to consider while designing a new measuring tool or instrument are content, concurrent, and predictive validity. Content validity is the estimate of how much a measure represents every single element of a construct [19]. It is a nonstatistical type of validity that involves 'the systematic examination' of the test content to determine whether it covers a representative sample of the behavior domain to be measured. A test has content validity built into it by a careful selection of which items to include [21]. Items are chosen so that they comply with the test specification that is drawn up through a thorough examination of the subject domain. By using a panel of experts to review the test specifications and the selections of items, the content validity of a test can be improved. The experts will be able to review the items and comment on whether the items cover a representative sample of the behavior domain [13].

Concurrent validity measures the test against a benchmark test and a high correlation indicates that the test has strong criterion validity [19]. For example, a new receptive vocabulary test might be correlated with the well-established Peabody Picture Vocabulary Test-Revised [22] to show the concurrent validity of the new test. A moderate, positive correlation is good for the new test. However, if the correlation is too high, there may be questions of the need for the new test [3].

Predictive validity is a measure of how well a test predicts abilities [19]. Predictive validity is also referred to as criterion validity. Broadly speaking, a criterion is any variable (e.g. language development) one wishes to explain and/or predict using information from other variables [13]. Predictive validity is the accuracy with which a test predicts future performance on a related task. It involves testing a group of participants for a certain construct and then comparing them with the results obtained at some point in the future. For example, a graduate student's score on comprehensive examinations might predict whether or not he or she will be a competent clinician. Thus, future performance is the criterion used to evaluate the predictive validity of a measure, the comprehensive examination in this case [3].

Reliability

Reliability refers to the consistency with which the same event is measured repeatedly. Scores are reliable if they are consistent across repeated testing or measurement. The concept of reliability applies to any kind of measure, including standardized tests [1].

Most measures of reliability are expressed in terms of correlation coefficient. The correlation coefficient is a number or index that indicates the relationship between two or more independent measures. It is usually expressed through the Pearson Product-Moment r (often referred to as Pearson's r). An r value of 0.00 indicates that there is a relationship between two measures. The highest possible positive value is 1.00. Conversely, the lowest possible negative value of r is -1.00 . The closer r is to 1.00, the greater the reliability of the test or the measurement. There are several types of reliability of a measure or a test.

Interobserver or interjudge (inter-rater) reliability refers to the extent to which two or more observers agree in measuring an event. For example, if three judges rate the fluency of a participant independently, there is a high interjudge reliability if there is good agreement between judges. Optimally, good agreement results in an interjudge reliability coefficient of 0.90 or more.

Intraobserver or intrajudge reliability refers to the extent to which the same observer repeatedly measures the same event consistently. For example, if the same clinician rates a child's intelligibility over several occasions, those ratings should be consistent if there is good intraobserver reliability (assuming that the child's intelligibility has not changed).

Alternate form reliability, also known as parallel form reliability, is based on the consistency of measures when two parallel forms of the same tests are administered to the same individuals. For example, the Test of Nonverbal-Intelligence-Third Edition (TONI-3) [23] includes form A and form B. If both these forms are administered to an adult client and the scores are very similar, then the TONI-3 has alternate form reliability.

Split-half reliability is a measure of the internal consistency of a test. Split-half reliability is determined by showing that the responses to items on the first half of a test are correlated with the responses given on the second half or the responses to even-numbered items should correlate with responses to odd-numbered items. Split-half reliability generally overestimates reliability because it does not measure the stability of scores over time.

Evidence-based practice

After examining the internal and external validity of research results, clinicians may choose techniques that are supported by well-designed and well-executed therapy efficacy studies [3]. Clinicians have to use techniques and procedures developed by research methodologies to consolidate, improve, develop, refine, and advance clinical aspects of their practice to serve their patients better [4]. Levels of evidence are often classified into three major classes: class I evidence (on the basis of a randomized group experimental design study) is the best evidence supporting a procedure. Class II evidence is based on well-designed studies that compare the performance of groups that are not randomly selected or assigned to different groups. Class III evidence is based on expert opinion and case studies. This is the weakest of the levels of evidence [3].

An alternative way of classifying evidence for clinical procedures accepts all valid research designs and is based on research that is uncontrolled, controlled, and replicated by the same or different investigators. This hierarchy of evidence moves from the least desirable to the most desirable evidence [1]:

- Level 1. Expert advocacy: There is no evidence supporting a treatment; the procedure is advocated by an expert.
- Level 2. Uncontrolled unreplicated evidence: A case study with no control group and the research was carried out once.
- Level 3. Uncontrolled directly replicated evidence: The study did not involve a control group but was repeated by the same researcher in the same setting and has obtained the same or similar levels of improvement.
- Level 4. Uncontrolled systematically replicated evidence: The study did not involve a control group but was repeated by another researcher in another setting with different patients and has obtained the same or similar levels of improvement.
- Level 5. Controlled unreplicated evidence: This is the first level at which efficacy is substantiated for a treatment procedure.

Level 6. Controlled directly replicated evidence: The study that involves a control group was repeated by the same researcher in the same setting and has obtained the same or similar levels of improvement. The technique is now known to produce the same effects, at least in the same setting.

Level 7. Controlled systematically replicated evidence: This is the highest level of evidence. The study involves a control group and was repeated by another researcher in another setting and has obtained the same or similar levels of improvement. This shows that the technique studied will produce the same effect under varied conditions. A technique that reaches this level may be recommended for general practice.

A critical examination of research evidence is at the heart of evidence-based practice. Clinicians should choose techniques that are supported by well-designed and well-executed treatment efficacy studies [3].

Conclusion and recommendations

(a) Research should be considered an integral part of any clinical practice. (b) Researchers have to keep the concept of validity, reliability, and evidence-based practice in mind at all times when designing a study. (c) A good researcher will discuss the project design with an advisor or a group of colleagues to help ensure that validity is preserved at every stage of the process. (d) Research should be considered as an integral part of any clinical practice. A researcher must think very carefully about the population that will be included in the study and how to sample that population. (e) Correctly forming a research problem, sticking to research ethics while carrying out research work, and closely reviewing the selected literature before as well as during the study. (f) Randomly selecting a representative sample of adequate size to represent the targeted population. (g) Identifying variables before starting research, choosing appropriate study designs, and appropriately dealing with missing data and confounding (extraneous) variables.

Acknowledgements

Conflicts of interest

There are no conflicts of interest.

References

- 1 PRO-ED. *Clinical research in communication disorders. Principles and strategies*. 3rd ed. Austin, TX: Pro-ed Inc.
- 2 Maxwell DL, Satake E. *Research and statistical methods in communication disorders*. 1st ed. Baltimore: Lippincott Williams & Wilkins; 1997.
- 3 Mckibbin CR, Hedge MN. *An advanced review of speech and language pathology*. 2nd ed. United States of America: Pro-ed Inc.; 2006.
- 4 Kumar R. *Research methodology: a step-by-step guide for beginners*. 3rd ed. Thousand Oaks, CA: Sage Publication limited; 2011.
- 5 Trochim WM. 'Probability sampling' research methods knowledge base. 2nd ed. <http://www.socialresearchmethods.net/kb/samprob.php>; 2009.
- 6 Pedhazur EJ. *Measurement, design and analysis: an integrated approach*. 1st ed. Hillsdale, NJ: Psychology Press; 1991.

- 7 Cook TD, Campbell DT. Quasi-experimentation: design and analysis issues for field settings. *Chiago*. Houghton Mifflin; 1979.
- 8 Petrie A, Caroline S. *Medical statistics at a glance*. 3rd ed Singapore: John Wiley & Sons Ltd Publication; 2009.
- 9 Patton MQ. *Qualitative evaluation and research methods*. 2nd ed. Newbury Park, CA: Sage; 1990.
- 10 Schiavetti N, Metz DE. *Evaluating research in communicative disorders*. 4th ed. Needham Heights, MA: Allyn and Bacon; 2002.
- 11 Weaver A, Goldberg S. *Clinical biostatistics made ridiculously simple*. 1 ed. Miami: Medmaster Inc. Miami; 2011.
- 12 Faragher B. *Essential statistics for medical examinations, PasTest*. 2nd revised ed. Knuthford, England; 2005.
- 13 Pelham BW, Blanton H. *Conducting research in psychology: measuring the weight of smoke*. 3rd ed. Belmont, CA: Wadsworth Publishing; 2006.
- 14 Stephen BH, Steven RC, Warren SB, Deborah GG, Thomas BN, Bulger RE. *Designing clinical research*. 3rd ed. USA: Lippincott Williams & Wilkins; 2007.
- 15 Creswell JW. *Educational research: planning, conducting and evaluation quantitative and qualitative research*. 3rd ed. Upper Saddle River, NJ: Pearson Education; 2008.
- 16 Pelham BW. *Conducting research in psychology*. 3rd ed. Belmont, CA: Wadsworth Publishing; 2006.
- 17 Silverman FH. *Research design and evaluation in speech-language pathology and audiology*. 4th ed Boston: Allyn and Bacon; 1997.
- 18 William LH. *Statistics*. Orlando, FL, USA: Saunders College Publishing; 1988.
- 19 Martyn S. Types of validity; 2009. Available at: <http://www.experiment-resources.com/types-of-validity.html>. [Accessed 31 January 2012].
- 20 Campbell DT, Stanely JC. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally; 1996.
- 21 Anastasi A, Urbina S. *Psychological testing*. 7th ed. Upper Saddle River, NJ: Prentice Hall; 1997.
- 22 Dunn LM, Dunn LM. *Peabody picture vocabulary test-revised*. Circle Pines: American Guidance Service; 1981.
- 23 Brown L, Sherbenou RJ, Johnson SK. *Test of non-verbal intelligence*. 3rd ed. Austin, TX: Pro-ed Inc; 1997.