

## Tutorial

# Statistical, Practical, Clinical, and Personal Significance: Definitions and Applications in Speech-Language Pathology

Anne K. Bothe<sup>a</sup> and Jessica D. Richardson<sup>b</sup>

**Purpose:** To discuss constructs and methods related to assessing the magnitude and the meaning of clinical outcomes, with a focus on applications in speech-language pathology.

**Method:** Professionals in medicine, allied health, psychology, education, and many other fields have long been concerned with issues referred to variously as practical significance, clinical significance, social validity, patient satisfaction, treatment effectiveness, or the meaningfulness or importance of beyond-clinic or real-world treatment outcomes. Existing literature addressing these issues from multiple disciplines was reviewed and synthesized.

**Conclusions:** *Practical significance*, an adjunct to *statistical significance*, refers to the magnitude of a change or a differ-

ence between groups. The appropriate existing term for the interpretation of treatment outcomes, or the attribution of meaning or value to treatment outcomes, is *clinical significance*. To further distinguish between important constructs, the authors suggest incorporating as definitive the existing notion that *clinical significance* may refer to measures selected or interpreted by professionals or with respect to groups of clients. The term *personal significance* is introduced to refer to goals, variables, measures, and changes that are of demonstrated value to individual clients.

**Key Words:** clinical significance, treatment outcomes, efficacy

Research and practice in speech-language pathology draw on several traditions, including constructs and methods first developed in areas ranging from medicine to psychology to education. One of the many concerns shared across these disciplines, and many others, involves the results of interventions that are intended to help a patient, client, or learner with a disorder, problem, condition, or complaint. Put simply, interventions are supposed to help. Assessed more thoroughly, this initially simple idea leads quickly to multiple interacting constructs that have been referred to across various traditions as outcomes assessment, outcomes measurement, practical significance, clinical significance, social validity, patient-reported outcomes, and other related terms.

The purpose of this article was to review several issues related to the application of these ideas to clinical research and practice in speech-language pathology. More specifically, the following sections (a) describe the constructs that have

previously been referred to as statistical, practical, and clinical significance; (b) differentiate among them and relate them to other relevant concepts and terminologies; and (c) provide several examples of how these concepts have been used, and could better be used, to improve clinical research and clinical practice in speech-language pathology. Most importantly, this article introduces the term *personal significance* to label a construct that has been widely recognized but not previously named and that represents, in one view, the single most important goal of clinical research or practice in any health care discipline or helping profession.

## Statistical Significance: Do the Groups Differ?

In a typical group-design treatment experiment, the members of at least one group of participants are measured, provided with (or purposely not provided with) one or more interventions, and then measured again (see Meline, 2010, or Schiavetti, Metz, & Orlikoff, 2011, for discussions of basic group-design research methods as applied to speech-language pathology). The groups' results are compared, usually through statistical tests of differences between means (e.g., *t* tests or analyses of variance). If differences are identified, according to the standard logic and procedures of inferential statistics, then the treatment is said to have resulted in a statistically significant difference.

Many philosophical and methodological weaknesses of this basic approach to treatment research have been discussed

<sup>a</sup>University of Georgia, Athens

<sup>b</sup>University of South Carolina, Columbia

Correspondence to Anne K. Bothe: abothe@uga.edu

Editor and Associate Editor: Laura Justice

Received April 15, 2010

Revision received September 22, 2010

Accepted March 25, 2011

DOI: 10.1044/1058-0360(2011/10-0034)

for decades, and alternatives have been proposed from several different points of view (Altman et al., 2001; Bakan, 1966; Barlow & Hersen, 1984; Carver, 1978; Cohen, 1988; Freeman, deMarrais, Preissle, Rouston, & St. Pierre, 2007; Gardner & Altman, 1986; Hugdahl & Ost, 1981; J. C. Ingham, 1990; Kazdin, 1982; Morrison & Henkel, 1970; B. Thompson, 2002; Tukey, 1991). Even within the traditions that accept group-design experiments as central to the development of science-based intervention practices, problems with depending on statistical significance testing to evaluate research results have been widely acknowledged. Indeed, review articles in this area (e.g., Kirk, 1996; B. Thompson, 2002) refer to articles published almost 100 years ago (Berkson, 1938; Boring, 1919) to support their point that “criticisms of statistical testing are virtually as old as the method” (B. Thompson, 2002, p. 64). Clinically and conceptually, the main problem does not lie with the group design itself; indeed, a well-controlled and carefully conducted group study can provide very strong, and very necessary, evidence about the typical effects of an intervention for a group of people. An important problem arises, however, when too much emphasis, or inappropriate emphasis, is placed on statistical significance testing as the method for assessing differences between pretreatment and posttreatment group mean data, or between treated and untreated groups. As all introductory statistical textbooks repeat, tests of statistical significance provide information only about how unlikely the obtained difference between the samples would have been if the populations from which the samples were drawn were actually identical (e.g., Sloan, Symonds, Vargas-Chanes, & Fridley, 2003). Tests of statistical significance are not intended to, and do not, provide information about whether an obtained difference between groups was large, helpful, curative, restorative, or important in any way other than being relatively unlikely to have occurred in the absence of an underlying difference in the populations. In practice, for clinical research, the long-recognized issue is that a statistically significant difference can be obtained between a group who received treatment and a group who did not receive treatment even if the change or the difference was small and even if the people who received the treatment still have noticeable problems. To address questions about the amount or the meaning of change, therefore, measures of statistical significance alone have long been recognized to be insufficient; measures of *practical significance* or *clinical significance* are required to address these questions.

### **Practical Significance: How Much Do the Groups Differ?**

Practical significance refers to the size of the difference between groups or occasions, or the extent to which two or more distributions of scores can be said to differ. The best known measures of practical significance are probably the family of related calculations known collectively as measures of effect size (Cohen, 1988; Keppel & Wickens, 2004; Kirk, 1996; B. Thompson, 2002; Young, 1993). The most common may be Cohen’s *d*, defined as the difference between the means of two distributions divided by the pooled

standard deviation of the two distributions (Cohen, 1988). (Common variations use a control standard deviation or the standard deviation of the pretreatment distribution; see Keppel & Wickens, 2004; Meline & Schmitt, 1997; Sloan et al., 2003.) The resulting number is unitless but may be interpreted using the standard deviation as a unit of measure: If the pretreatment standard deviation was used in the calculation, then an effect size of 0.5 is interpreted as showing that the mean posttreatment score is 0.5 pretreatment standard deviations away from the mean pretreatment score. Like any other statistic, effect sizes can also be interpreted not as point values but in terms of a range of more likely values (a confidence interval). In the absence of any other empirical information, the point values 0.2, 0.5, and 0.8 have been referred to as small, moderate, and large effect sizes, respectively (Cohen, 1988; Keppel & Wickens, 2004), although the widespread use of such generic values conflicts with Cohen’s (1988) original intent and has attracted considerable criticism (Cohen, 1992; Cortina & Landis, 2009; Lenth, 2001; B. Thompson, 2002).

The reporting of effect sizes has increasingly become recommended and even required in many social science and health sciences journals (American Psychological Association, 2001, 2009; B. Thompson, 2002). As useful as the information they provide may be, however, it is also important to remember that effect sizes are simply adjuncts to tests of statistical significance. They should not be overinterpreted, for several reasons. First, a *d* of 1.0, for example, which would traditionally be interpreted as quite large, suggests only (depending on the specific formula chosen) that the second distribution’s mean has shifted by an amount equal to 1 *SD* of the original pooled distribution. In such a case, much of the two distributions may still overlap. Secondly, even though effect sizes are more robust to sample size issues than are many other statistics, sample size does remain a concern; the use of confidence intervals, rather than point estimates, can begin to address this problem, but this additional step appears to be uncommon.

Finally, the most important issue may be that a measure of effect size cannot and does not address whether an obtained difference reflects a change that any person would describe as a good change, an important change, a big change given the entire range of possible scores, or enough of a change to alter how the members of that group feel or function in their own daily lives. This interpretation or attribution of meaning falls within the purview of another construct: clinical significance.

### **Clinical Significance: What Does the Difference Between Groups Mean?**

The term *clinical significance* is used in many ways, but the definitive element in most scholarly definitions of the term is that it refers to a value placed upon a result by a person. Thus, clinical significance implies that—and should be used when—a researcher, practitioner, client, or other person is attempting to interpret, ascribe meaning to, or make judgments about the value of a test result, a matter of physical status, or a change or difference between groups.

Accordingly, in the context of treatment research and clinical interventions, clinical significance has been defined as whether an intervention “makes a real (e.g., genuine ... noticeable) difference in everyday life to the clients or to others with whom the clients interact” (Kazdin, 1999, p. 332) and in terms of a “recognizable treatment change that is valued by the clinician, client, and relevant others” (Finn, 2003, p. 215).

Understanding the formal definitions and the informal uses of the term *clinical significance*, as well as the distinction between practical significance and clinical significance, is complicated by several factors. First, some definitions of clinical significance use the word “practical” to refer not to practical significance (as discussed in the previous section) but to clinical significance. Houle and Stump (2008, p. 5), for example, defined clinical significance as “the practical or applied value of a treatment effect”; Johnson (2006, p. 21) referred to the “clinical or practical significance of results” as one construct. Others have resorted to somewhat awkward metaphors to define clinical significance (e.g., “palpable benefits”; Kazdin & Kendall, 1998) or to the presumably unintended implication in defining practical or clinical significance as “meaningful change” or “important change” that statistically significant differences between groups of scores are not meaningful or important. Statistically significant differences between groups of scores are both meaningful and important, as are effect sizes; the problem is only that they do not mean what some clients or clinicians would like them to mean or have assumed them to mean. Despite all these terminological complexities, however, practical significance and clinical significance can be straightforwardly differentiated by recognizing that practical significance refers to the magnitude of a change (Kirk, 2001; Nickerson, 2000), whereas clinical significance refers to the meaning of a change (Ogles, Lunnen, & Bonesteel, 2001).

Many methods for measuring clinical significance have been proposed and used, and several thoughtful and comprehensive discussions of their history and development are available (see Follette & Callaghan, 2001; Kazdin, 1999; Ogles et al., 2001). One of the more influential attempts to formalize the measurement of clinically significant change (Jacobson, Follette, & Revenstorf, 1984) defined this construct as change that (a) is in excess of that which could be expected by measurement error alone and (b) represents movement from a dysfunctional range to a normal or functional range on the measure of interest. Based on the work of Nunnally and Kotsche (1983), Jacobson et al. (1984) addressed the first of these criteria by defining a reliable change index (RC) as the difference between pretest and posttest scores divided by the standard error of measurement for pretest scores. Correcting the change or difference score by the standard error of pretest scores was intended to allow the determination that an obtained change in scores “represents more than the fluctuations of an imprecise instrument” (Jacobson et al., 1984, p. 344). Christensen and Mendoza (1986) subsequently argued that correcting by the standard error of the difference, rather than the standard error of measurement for pretest scores alone, provides a better estimate of the construct in question, creating a variation on Jacobson et al.’s original index that is defined as:  $RC = (\text{posttest}$

score – pretest score)/ $SE_{diff}$ , where  $SE_{diff}$  is the standard error of the difference.  $SE_{diff} = (SD_b)(2)^{1/2}(1 - r)^{1/2}$ , where  $SD_b$  is the standard deviation of the baseline observation, and  $r$  is the reliability of the measurement (see Beal & Duckro, 2003; Evans, Margison, & Barkham, 1998; Jacobson & Truax, 1991). If the obtained RC is greater than a selected critical  $z$  value (e.g., 1.96 to approximate a 95% confidence level), then the conclusion is drawn that a reliable change has occurred. Later authors corrected these formulas again, based on their concerns that a simple subtraction of pretest scores from posttest scores does not take regression toward the mean into account (see Peterson, 2008, for an accessible review of this and related issues and formulas).

Regardless of the specific formula selected, all variations on an RC show only that change has occurred. For that change to be deemed clinically significant, in Jacobson et al.’s terms, it must also be interpreted as representing movement from a “dysfunctional” group to a “functional” group (E. M. Anderson & Lambert, 2001; Beal & Duckro, 2003; Jacobson & Truax, 1991). How to identify an appropriate cutoff point between or within the “functional” and “dysfunctional” distributions is widely recognized as a complex question, and several authors have provided elegant mathematical models and simulations in attempting to answer it (see Jacobson, Roberts, Berns, & McGlinchey, 1999). This body of work also recognizes that requiring a return to normal functioning may not always be realistic, for any number of reasons (Jacobson et al., 1984; Kazdin, 1977). There are strong arguments to be made for keeping that goal in mind; if nothing else, as Jacobson et al. (1999, p. 300) phrased it, “Consumers enter therapy expecting that their presenting problems will be solved.” Nevertheless, in recognition of clinical realities, these issues have led to several modifications to the original presentations of clinically significant reliable change. Therefore, reliable change in addition to movement within or into a normal range may be referred to by some authors as *recovered*, reliable change in the direction of the normal range as *improved*, no reliable change (including change that does not exceed that which could have occurred through measurement error alone) as *unchanged*, and finally, reliable change in the direction opposite of the normal range as *deteriorated* (Beal & Duckro, 2003; Hageman & Arrindell, 1999; Tingey, Lambert, Burlingame, & Hansen, 1996).

Several points are worth emphasizing about all of these variations on Jacobson et al.’s (1984) original formula and definitions of clinically significant change. First, they are distributional or group-level measures. That is, they may address change on the part of individuals, but they do so with respect to distributions, or groups, of scores. Two of the guiding assumptions behind such procedures, as many of the authors in this area have discussed, are that (a) the distributions of scores in the affected and unaffected populations are known, and (b) the scores in question represent relevant measurements. The latter of these two ideas is the more conceptually difficult and leads to an intriguing paradox: Reliable clinically significant change, as those terms are used by Jacobson et al. and in related calculations, may not be at all clinically significant, when that term is used more generally to mean that a change judged to be useful or important has occurred in how clients feel or function. This problem

also exists for less formal attempts to measure or describe clinically significant change. In sum, we do not need to know only that clients' scores have changed enough to justify their classification as "closer to normal" or as "performing much better" on some measure; we need to know that the clients themselves find the improvements to be helpful. Fayers and Machin (2007, p. 17) stated this point delightfully succinctly, in the single subheading within their section labeled "How to measure quality of life": "Ask the patient."

### **Personal Significance: Did This Change Solve a Person's Problem?**

As discussed in preceding paragraphs, the existing term *clinical significance* commonly refers to discussions about whether reliable and/or important change has occurred that is sufficient to change the professional's clinical description of, or clinical label for, a client or a group of clients. Jacobson et al.'s work attempted to quantify this idea, but the same notion is present even in less formal applications (i.e., usages that equate clinical significance with a clinician-defined or researcher-defined amount of change). Thus, clinical significance may address the need for beyond-clinic or real-world outcomes, but it does so through group-level or distributional measures that often have, at their core, the researcher's or the clinician's decisions or value judgments about the importance or meaningfulness of treatment results. Professionals' informed judgments are important, and we suggest recognizing and, indeed, incorporating as definitive this professional-level aspect of clinical significance.

A very different construct, and the critical remaining need, involves that which we here propose to call *personal significance*. This term refers to whether individual clients report feeling improvements that matter to them in the context of their own lives and, more importantly, whether they demonstrate functioning in ways that reflect improvements. Personal significance, like clinical significance, attempts to address the value of a change to a human being, and the construct that we have labeled personal significance has certainly been incorporated into many previous discussions of clinical significance. The two may be usefully distinguished, however, by the fact that clinical significance reflects or may incorporate the clinician's judgments, decisions, or values, whereas personal significance reflects the client's.

Both clinical significance and personal significance are important to a complete description and understanding of clinical change, and both have been addressed (using the label *clinical significance*) in previous literature. In psychology, for example, Battle et al. (1966) introduced the Target Complaints model for eliciting clients' descriptions of their own problems, because "the alleviation of these complaints is the criterion of the efficacy of the treatment" (p. 185). Baer (1988) developed similar ideas 20 years later with his emphasis on the need to treat the client's point of complaint. Strupp and Hadley (1977) included client-based measures in the context of measuring clinical outcomes from the point of view of the individual, the health professional, and society. Kendall and Norton-Ford (1982), similarly, referred to measures at both the "specifying level" and the "general impact

level." Jacobson et al. (1999) noted, as quoted in the preceding section, that clients present with a self-defined problem and generally expect the professional to solve that problem, not a different problem. Rosen and Proctor's (1981) differentiation among intermediate, instrumental, and ultimate outcomes reflects some of the same issues. Other related alternatives, also rooted primarily in psychology and education, include attempts to ensure that treatments have what other persons will recognize as socially valid outcomes, or outcomes that are perceived as improvements by observers from the clients' own real-world contexts (Foster & Mash, 1999; Kazdin, 1977; Wolf, 1978).

In the health outcomes literature, similar constructs are phrased in terms of the need to measure, and to lessen, the burden that health conditions place on patients, clients, and their families (M. S. Thompson, 1986) and, from other points of view, on care providers and systems (the original use of the term; e.g., Black & Pole, 1975), and on society as a whole (Murray, 1994). This idea is also reflected in discussions of the relative merits of clinician-reported outcomes and patient-reported outcomes (Cella et al., 2007; Wiklund, 2004; see also [www.nihpromis.org/default.aspx](http://www.nihpromis.org/default.aspx)), and in approaches to health care that focus not only on clinical signs and symptoms but also on clients' self-reported health-related quality of life (Fitzpatrick et al., 1992; McHorney & Tarlov, 1995; Walker & Rosser, 1993). Many of the same issues are addressed by the constructs of client or consumer satisfaction (Donabedian, 1988), clinimetric approaches to assessing the importance of variables (known as the determinant attribute approach in other traditions; see W. T. Anderson, Fox, & Fulcher, 1976; Franic, Haddock, Tucker, & Wooten, 2008; Juniper, Guyatt, Streiner, & King, 1997), anchor-based attempts to identify the client-centered minimal clinically important difference (Jaeschke, Singer, & Guyatt, 1989; Lauridsen, Hartvigsen, Manniche, Korsholm, & Grunnet-Nilsson, 2006; Sloan et al., 2003), discussions of accountability in health care (Emanuel & Emanuel, 1996), and many others. The World Health Organization's older differentiations among impairment, disability, and handicap (World Health Organization, 1980) and its current differentiations among function, structure, activities, and participation (World Health Organization, 2001) also address related ideas; the disorder itself is separated from the influence of that disorder on the person living with it and on other persons and systems.

All of these ideas recognize that the person best suited to judge the meaningfulness of a change caused by a putatively therapeutic intervention may be the person to whom those changes were meant to occur. In other words, many different disciplines recognize that a complete clinical measurement system should include both group-level or clinician-identified measures and also measures that are defined or interpreted at the level of individuals, their perceptions of their own lives, or the changes they expected or have experienced as a result of treatment. There remains an important problem of nomenclature, however, because the phrase *clinical significance* has been used to refer to all of these related but distinct concepts and calculations: formulas about changes in distributions of scores, decisions about whether a change was large enough to warrant a different diagnostic label, general discussions of

the need for treatments to produce real-life or real-world changes, and even more general discussions of whether this change was what the client wanted. Referring to the client's needs and judgments as part of a construct known as *personal significance*, while reserving the term *clinical significance* for clinicians' judgments or group-oriented decisions, serves to differentiate clearly the client's view from the clinician's. It also serves to emphasize the individual's self-judged progress as distinct from comparisons between groups. Again, both clinical significance and personal significance are important; the issue is that neither can be fully understood while the two remain intertwined. We propose and recommend a clear distinction between the two.

## Applications to Research and Practice in Speech-Language Pathology

### *Historical and Recent Usage of Significance Terms*

Within the speech-language pathology literature, methodologists have emphasized for many years that statistical significance does not imply practical or clinical significance (e.g., Young, 1976, 1993). Despite those recommendations, publications in the journals of the American Speech-Language-Hearing Association (ASHA) tended historically to report only statistical significance (Meline & Schmitt, 1997), although this pattern has changed during approximately the last 10 years (Meline & Wang, 2004). Indeed, of 22 articles that reported statistical significance measures in the two most recent complete volumes of three ASHA journals (the *American Journal of Speech-Language Pathology*, the *Journal of Speech, Language, and Hearing Research*, and *Language, Speech, and Hearing Services in Schools*), 16 (more than 72%) were identified by the present authors as having included a measure of practical significance (most commonly a variation on Cohen's *d*).

This change toward the reporting of practical significance measures in speech-language pathology research is a positive change, but a more complex problem emerges when the reporting of effect sizes seems to be implicitly equated with the measurement, or even the establishment, of clinical significance (Meline & Paradiso, 2003; Meline & Schmitt, 1997; Schuele & Justice, 2006). Bain and Dollaghan's (1991) discussion of what they referred to as three overarching criteria for interpreting clinical change, for example, has become relatively widely known among speech-language pathology treatment researchers, but Bain and Dollaghan defined clinical change in terms of research quality (internal validity of research designs) and the magnitude of the change (statistical significance and practical significance, as effect size), not in terms of clinical significance (the meaningfulness of outcomes to persons). Many reports of treatment research in speech-language pathology also imply that effect sizes can measure or establish the clinical significance, clinical meaningfulness, or clinical importance of their findings. Apel and Masterson (2001) and Friel-Patti, DesBarres, and Thibodeau (2001), for example, both defined an effect size of 1.0 as clinically significant. Gillam, Crofford, Gale, and Hoffman (2001) and Swanson, Fey, Mills, and Hood (2005), similarly, described change as clinically significant if posttest

scores were outside the 95% confidence intervals of pretest scores, and other authors have defined clinically significant change as requiring an increase of a certain number of points on a specific instrument (Elman & Bernstein-Ellis, 1999; Katz & Wertz, 1997; Walker-Batson et al., 2001). These approaches represent important attempts to move beyond merely reporting statistically significant differences toward identifying changes that are large enough that they might be assumed to be meaningful. They remain somewhat problematic, however, in that they blur the distinction between the magnitude of a change and the meaningfulness of that change. Despite the apparent face validity of the assumption that a larger change is more important or more functionally meaningful, it is known to be inaccurate; the relationship between amount of change and perceived meaningfulness, usefulness, or importance of change is not a simple one (Copay, Subach, Glassman, Polly, & Schuler, 2007; Kazdin, 1999; Mintz, 1972; Ogles et al., 2001).

Other efforts in speech-language pathology have attempted to address the meaningfulness of treatment outcomes in ways other than focusing on the magnitude of change. Researchers and treatment providers who work with persons with aphasia, for example, have long been encouraged to view aphasia treatment from a client-centered perspective that emphasizes communicative effectiveness in real-life situations (e.g., Brookshire, 1986; Rosenbek, LaPointe, & Wertz, 1989). Aphasia is also distinguished by the existence of widely used, standardized instruments designed to measure abilities that the test developers intended to be relevant in everyday life (e.g., the Communication Activities of Daily Living, Second Edition; Holland, Frattali, & Fromm, 1999); this instrument was intended to serve, in other words, as a measure of clinical significance (meaningfulness of abilities or changes, as defined and assessed by the clinician). Client-centered ideas for aphasia treatments were also formalized in 2000 as a consensus statement supporting a Life Participation Approach to Aphasia (LPAA; Kagan & LeBlanc, 2002; LPAA Project Group, 2001), which emphasizes "empower[ing] the consumer to select and participate in the recovery process and to collaborate on the design of interventions" (LPAA Project Group, 2001, p. 235)—that is, the LPAA emphasized a decade ago some parts of the construct that we propose to call personal significance.

Similarly, some of the earliest stuttering-specific measurement instruments were intended to identify the client's specific complaints with respect to avoidance or other reactions, so that these problems could be addressed in counseling or other treatment (Shumak, 1955; Woolf, 1967). Measurement procedures focused more directly on the speech of clients who stutter have also suggested that clients should self-select the beyond-clinic speaking situations within which speech will be recorded and measured, again so that problems as identified by the client, rather than by the clinician, can serve as the focus of treatment (Costello & Ingham, 1984; R. J. Ingham & Cordes, 1997). Both of these examples might be described, in the present terms, as attempts to identify treatment goals and outcomes that will be of personal significance to individual clients.

There have been, in summary, many previous attempts in speech-language pathology to demonstrate, or at least to

consider, the statistical, practical, clinical, and personal significance of treatment outcomes. The need, moving forward, will be to build on these previous attempts and previous successes within the context of our long-standing commitment, as professionals, to hold paramount the welfare of the clients we serve (ASHA, 2010). Such a focus on the client-centered personal significance of our methods and measures is consistent with our ethical precepts as a profession, and it is also consistent with currently dominant frameworks for practice (e.g., evidence-based practice emphasizes the importance of the client's preferences and perspectives in addition to research evidence and the clinician's expertise; Bothe, 2003; Straus, Richardson, Glasziou, & Haynes, 2005).

### **Recommendations for the Future**

Given the issues discussed throughout this article, it would appear to be of some value for clinical researchers, in particular, to consider specifying which types of significance their studies are attempting to address. We do not mean to imply, by differentiating among statistical, practical, clinical, and personal significance, that all studies must address all four. Group-design studies that demonstrate a statistically significant difference between two different treatments, or a practically significant difference on a given measure from pretreatment to posttreatment, certainly have their place, especially if the effect size is relatively high and with a relatively narrow confidence interval. Problems emerge only when a statistically or practically significant result is interpreted as if it were necessarily of clinical or personal significance; a focus on one or two types of significance, if that focus is explicit and recognized, could represent a strength, not a weakness, of clinical research.

On a related note, however, we question those approaches to clinical trials research that focus too closely on the recommendation to select only one primary outcome variable (Altman et al., 2001; Onslow, Jones, O'Brian, Menzies, & Packman, 2008). This recommendation emerged from the need to prevent the spurious findings of statistical significance that can occur as a function of conducting multiple tests, or as a result of analyzing correlated or nonindependent measurements, among other reasons. Such statistical concerns should not be allowed to detract from the larger conceptual need for treatment research to address the multiple interacting variables that may define the real-life value of treatment outcomes or their clinical significance and personal significance (Wade, Smeets, & Verbunt, 2010). From this point of view, then, a more general recommendation might be that clinical research should attempt, where reasonable, to address the well-defined clinical and personal significance of its methods and its outcomes. In some instances, this goal will be achievable through a single, well-selected measure; in other circumstances, clinical research will need more than one outcome variable to be able to measure, and to demonstrate, all of the changes that are important to professionals, families, clients, and other stakeholders.

It is also clear that differentiating between clinical significance and personal significance will require the development and the careful use of validated measurement

methods and materials that are appropriate to each construct. Methodologically, the label *personal significance* is not intended to refer necessarily to self-reported data or patient-reported outcomes. The designation of patient-reported is a description of measurement method, not a description of meaning or importance. Many questionnaire items, for example, can be perceived by the client as irrelevant, meaningless, or built on the researcher's point of view rather than on the client's needs, and the resulting data can be patient-reported without being personally significant. Equally, it must be noted that measures of attitudes, emotions, private details of interpersonal relationships, or any other ostensibly personal topic are not necessarily of personal significance in any given case; they would only be so if the client's complaint were about his or her attitudes, emotions, or other specific topic. Similarly, personal significance cannot be addressed through long surveys that include long lists of many possible problems previously identified by other clients or other professionals (see Franic & Bothe, 2008). Instead, the notion of personal significance suggests a focused approach to outcomes measurement and a need for the use of individually identified and individually functional measures. Physiological, laboratory, and observer-generated measures can therefore be directly relevant and indeed critical to measuring personal significance, if those data are the means through which it can be demonstrated that the problem a client wanted solved has been addressed.

Measuring the personal significance of treatment outcomes, therefore, would seem to require one of two approaches, both of which begin, as Fayers and Machin (2007) recommended, with asking the client. First, one can ask clients to rate whether, or to what extent, they perceive an obtained treatment outcome to be important, meaningful, helpful, or personally significant. The means by which such questions are asked could take any one of many forms, although the measurement of personal significance must be differentiated from the administration of client satisfaction questionnaires; the tendency for most such questionnaires to produce spuriously high estimates of client satisfaction is widely known in business, marketing, and health care administration (see W. T. Anderson et al., 1976; Ankuta & Abeles, 1993; Ford, Bach, & Fottler, 1997; Mazor, Clauser, Field, Yood, & Gurwitz, 2002). To the extent that future thoughtful attempts to measure client satisfaction in speech-language pathology could be relevant to measuring the personal significance of treatment outcomes, treatment researchers might do well to begin with the classic texts and existing efforts in other disciplines (W. T. Anderson et al., 1976; Ankuta & Abeles, 1993; Cleary & McNeil, 1988) and then take advantage of existing clinimetric (i.e., determinant attribute) methodologies (Franic et al., 2008).

An equally reasonable approach to measuring personal significance and one that avoids the known problems with asking clients to rate their satisfaction with products or services might begin by asking clients which variables would need to improve, and by how much, for them to feel that personally significant outcomes have been achieved. The variables thus nominated, selected by individual clients as the most important to their own views of meaningful

improvements, would then need to be measured in ways that result in objective, professionally defensible data of demonstrable reliability, validity, and other important psychometric features (Beutler & Hamblin, 1986; Cordes, 1994; Kearns & Simmons, 1988; McCauley & Swisher, 1984; McHorney & Tarlov, 1995; Shriberg & Lof, 1991).

In conclusion, this article has introduced the term *personal significance* as a complement to the existing terms *statistical significance*, *practical significance*, and *clinical significance*. Of course, many of the complexities about how personalized measures can be usefully, reliably, and validly identified, measured, compared, and interpreted have yet to be addressed; some clients are more aware than others, for example, of their values, preferences, and needs with respect to their treatment plans or goals. If those and other complexities can be solved, however, then it may emerge that a combination of self-ratings of the personal significance of treatment-induced changes, plus professionally objective and professionally defensible data about variables selected by individual clients and showing that clients' concerns have been resolved, might represent the ideal evidence-based, client-centered, and outcomes-focused result of clinical research and practice in speech-language pathology and related disciplines.

## Acknowledgments

An earlier version of this article was first developed while the second author was associated with the University of Georgia. Our thanks to Duska M. Franic for sharing her expertise in pharmacoeconomics, quality of life measurement, and related areas, and to Patrick Finn for helpful comments about multiple versions of the manuscript.

## References

- Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., ... for the CONSORT Group. (2001). The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. *Annals of Internal Medicine*, *134*, 663–694.
- American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.
- American Psychological Association. (2009). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- American Speech-Language-Hearing Association. (2010). *Code of ethics*. Available from [www.asha.org/policy](http://www.asha.org/policy).
- Anderson, E. M., & Lambert, M. J. (2001). A survival analysis of clinically significant change in outpatient psychotherapy. *Journal of Clinical Psychology*, *57*, 875–888.
- Anderson, W. T., Fox, E. P., & Fulcher, D. G. (1976). Bank selection decision and market segmentation. *Journal of Marketing*, *40*, 40–45.
- Ankuta, G. Y., & Abeles, N. (1993). Client satisfaction, clinical significance, and meaningful change in psychotherapy. *Professional Psychology: Research and Practice*, *24*, 70–74.
- Apel, K., & Masterson, J. J. (2001). Theory-guided spelling assessment and intervention: A case study. *Language, Speech, and Hearing Services in Schools*, *32*, 182–195.
- Baer, D. M. (1988). If you know why you're changing a behavior, you'll know when you've changed it enough. *Behavioral Assessment*, *10*, 219–223.
- Bain, B. A., & Dollaghan, C. A. (1991). Treatment efficacy: The notion of clinically significant change. *Language, Speech, and Hearing Services in Schools*, *22*, 264–270.
- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, *66*, 423–437.
- Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: Strategies for studying behavior change* (2nd ed.). New York, NY: Pergamon.
- Battle, C. C., Imber, S. D., Hoehn-Saric, R., Stone, A. R., Nash, K. R., & Frank, J. D. (1966). Target complaints as criteria of improvement. *American Journal of Psychotherapy*, *20*, 184–192.
- Beal, D., & Duckro, P. (2003). Empirically documenting clinically significant change in rational emotive behavior therapy. *Journal of Rational-Emotive and Cognitive-Behavior Therapy*, *21*(2), 75–88.
- Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association*, *33*, 526–536.
- Beutler, L. E., & Hamblin, D. L. (1986). Individualized outcome measures of internal change: Methodological considerations. *Journal of Consulting and Clinical Psychology*, *54*, 48–53.
- Black, D. A. K., & Pole, J. D. (1975). Priorities in biomedical research: Indices of burden. *British Journal of Preventative and Social Medicine*, *29*, 222–227.
- Boring, E. G. (1919). Mathematical vs. scientific importance. *Psychological Bulletin*, *16*, 335–338.
- Bothe, A. K. (2003). Evidence-based treatment of stuttering: V. The art of clinical practice and the future of clinical research. *Journal of Fluency Disorders*, *28*, 247–258.
- Brookshire, R. H. (1986). *An introduction to aphasia* (3rd ed.). Minneapolis, MN: BRK.
- Carver, R. (1978). The case against statistical significance testing. *Harvard Educational Review*, *48*, 378–399.
- Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., ... on behalf of the PROMIS cooperative group. (2007). The Patient Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group during its first two years. *Medical Care*, *45*(5), S3–S11.
- Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy*, *17*, 305–308.
- Cleary, P. D., & McNeil, B. J. (1988). Patient satisfaction as an indicator of quality care. *Inquiry: A Journal of Medical Care Organization, Provision and Financing*, *25*, 25–36.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.
- Copay, A. G., Subach, B. R., Glassman, S. D., Polly, D. W., Jr., & Schuler, T. C. (2007). Understanding the minimum clinically important difference: A review of concepts and methods. *The Spine Journal*, *7*, 541–546.
- Cordes, A. K. (1994). The reliability of observational data: I. Theories and methods for speech-language pathology. *Journal of Speech and Hearing Research*, *37*, 264–278.
- Cortina, J. M., & Landis, R. S. (2009). When small effect sizes tell a big story, and when large effect sizes don't. In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the*

- organizational and social sciences (pp. 287–308). New York, NY: Routledge.
- Costello, J. M., & Ingham, R. J.** (1984). Assessment strategies for stuttering. In R. F. Curlee & W. H. Perkins (Eds.), *Nature and treatment of stuttering: New directions* (pp. 303–333). San Diego, CA: College Hill Press.
- Donabedian, A.** (1988). The quality of care: How can it be assessed? *JAMA: The Journal of the American Medical Association*, *260*, 1743–1748.
- Elman, R. J., & Bernstein-Ellis, E.** (1999). The efficacy of group communication treatment in adults with chronic aphasia. *Journal of Speech, Language, and Hearing Research*, *42*, 411–419.
- Emanuel, E. J., & Emanuel, L. L.** (1996). What is accountability in health care? *Annals of Internal Medicine*, *124*, 229–239.
- Evans, C., Margison, F., & Barkham, M.** (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health*, *1*(3), 70–72.
- Fayers, P. M., & Machin, D.** (2007). *Quality of life: The assessment, analysis and interpretation of patient-reported outcomes* (2nd ed.). Chichester, England: Wiley.
- Finn, P.** (2003). Evidence-based treatment of stuttering: II. Clinical significance of behavioral stuttering treatments. *Journal of Fluency Disorders*, *28*, 209–218.
- Fitzpatrick, R., Fletcher, A., Gore, S., Jones, D., Spiegelhalter, D., & Cox, D.** (1992). Quality of life measures in health care: I. Applications and issues in assessment. *British Medical Journal*, *305*, 1074–1077.
- Follette, W. C., & Callaghan, G. M.** (2001). The evolution of clinical significance. *Clinical Psychology: Science and Practice*, *8*, 431–435.
- Ford, R. D., Bach, S. A., & Fottler, M. D.** (1997). Methods of measuring patient satisfaction in health care organizations. *Health Care Management Review*, *22*, 74–89.
- Foster, S. L., & Mash, E. J.** (1999). Assessing social validity in clinical treatment research: Issues and procedures. *Journal of Consulting and Clinical Psychology*, *67*, 308–319.
- Franic, D. M., & Bothe, A. K.** (2008). Psychometric evaluation of condition-specific instruments used to assess health-related quality of life, attitudes, and related constructs in stuttering. *American Journal of Speech-Language Pathology*, *17*, 60–80.
- Franic, D. M., Haddock, S. M., Tucker, L. T., & Wooten, N.** (2008). Pharmacy patronage: Identifying key factors in the decision-making process using the determinant attribute approach. *Journal of the American Pharmacists Association*, *48*, 71–85.
- Freeman, M., deMarras, K., Preissle, J., Roulston, K., & St. Pierre, E. A.** (2007). Standards of evidence in qualitative research: An incitement to discourse. *Educational Researcher*, *36*, 25–32.
- Friel-Patti, S., DesBarres, K., & Thibodeau, L.** (2001). Case studies of children using Fast ForWord. *American Journal of Speech-Language Pathology*, *10*, 203–215.
- Gardner, M. J., & Altman, D. G.** (1986). Confidence intervals rather than P values: Estimation rather than hypothesis testing. *British Medical Journal*, *292*, 746–750.
- Gillam, R. B., Crofford, J. A., Gale, M. A., & Hoffman, L. M.** (2001). Language change following computer-assisted language instruction with Fast ForWord or Laureate Learning Systems software. *American Journal of Speech-Language Pathology*, *10*, 231–247.
- Hageman, W. J., & Arrindell, W. A.** (1999). Establishing clinically significant change: Increment of precision and the distinction between individual and group level of analysis. *Behavioral Research Therapy*, *37*, 1169–1193.
- Holland, A. H., Frattali, C., & Fromm, D.** (1999). *Communication Activities of Daily Living, Second Edition*. Austin, TX: Pro-Ed.
- Houle, T. T., & Stump, D. A.** (2008). Statistical significance versus clinical significance. *Seminars in Cardiothoracic and Vascular Anesthesia*, *12*, 5–6.
- Hugdahl, K., & Ost, L.** (1981). On the difference between statistical and clinical significance. *Behavioral Assessment*, *3*, 289–295.
- Ingham, J. C.** (1990). Issues of treatment efficacy: Design and experimental control. In L. B. Olswang, C. K. Thompson, S. F. Warren, & N. J. Minghetti (Eds.), *Treatment efficacy research in communication disorders* (pp. 51–61). Rockville, MD: American Speech-Language-Hearing Foundation.
- Ingham, R. J., & Cordes, A. K.** (1997). Self-measurement and evaluating stuttering treatment efficacy. In R. F. Curlee & G. S. Siegel (Eds.), *Nature and treatment of stuttering: New directions* (Rev. ed., pp. 413–437). Boston, MA: Allyn & Bacon.
- Jacobson, N., Follette, W. C., & Revenstorff, D.** (1984). Psychotherapy outcome research: Methods of reporting variability and evaluating clinical significance. *Behavior Therapy*, *15*, 336–352.
- Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B.** (1999). Methods for defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting and Clinical Psychology*, *67*, 300–307.
- Jacobson, N. S., & Truax, P.** (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.
- Jaeschke, R., Singer, J., & Guyatt, G. H.** (1989). Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials*, *10*, 407–415.
- Johnson, C. J.** (2006). Getting started in evidence-based practice for childhood speech-language disorders. *American Journal of Speech-Language Pathology*, *15*, 20–35.
- Juniper, E. F., Guyatt, G. H., Streiner, D. L., & King, D. R.** (1997). Clinical impact versus factor analysis for quality of life questionnaire construction. *Journal of Clinical Epidemiology*, *50*, 233–238.
- Kagan, A., & LeBlanc, K.** (2002). Motivating for infrastructure change: Toward a communicatively accessible, participation-based stroke care system for all those affected by aphasia. *Journal of Communication Disorders*, *35*(2), 153–169.
- Katz, R. C., & Wertz, R. T.** (1997). The efficacy of computer-provided reading treatment for chronic aphasic adults. *Journal of Speech, Language, and Hearing Research*, *40*, 493–507.
- Kazdin, A. E.** (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, *1*, 427–452.
- Kazdin, A. E.** (1982). *Single-case research designs: Methods for clinical and applied settings*. New York, NY: Oxford.
- Kazdin, A. E.** (1999). The measurement and meanings of clinical significance. *Journal of Consulting and Clinical Psychology*, *67*, 332–339.
- Kazdin, A. E., & Kendall, P. C.** (1998). Current progress and future plans for developing effective treatments: Comments and perspectives. *Journal of Clinical Child Psychology*, *27*, 217–226.
- Kearns, K. P., & Simmons, N. N.** (1988). Interobserver reliability and perceptual ratings. *Journal of Speech and Hearing Research*, *31*, 131–136.
- Kendall, P. C., & Norton-Ford, J. D.** (1982). Therapy outcome research methods. In P. C. Kendall & J. N. Butcher (Eds.),

- Handbook of research methods in clinical psychology* (pp. 429–460). New York, NY: Wiley.
- Keppel, G., & Wickens, T. D.** (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Kirk, R.** (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement, 56*, 746–759.
- Kirk, R. E.** (2001). Promoting good statistical practices: Some suggestions. *Educational and Psychological Measurement, 61*, 213–218.
- Lauridsen, H. H., Hartvigsen, J., Manniche, C., Korsholm, L., & Grunnet-Nilsson, N.** (2006). Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. *BMC Musculoskeletal Disorders, 7*, 82. doi:10.1186/1471-2474-7-82.
- Lenth, R. V.** (2001). Some practical guidelines for effective sample size determination. *The American Statistician, 55*, 187–193.
- LPA Project Group.** (2001). Life participation approach to aphasia: A statement of values for the future. In R. Chapey (Ed.), *Language intervention strategies in aphasia and related neurogenic communication disorders* (4th ed., pp. 235–245). Baltimore, MD: Lippincott Williams & Wilkins.
- Mazor, K. M., Clauser, B. E., Field, T., Yood, R. A., & Gurwitz, J. H.** (2002). A demonstration of the impact of response bias on the results of patient satisfaction surveys. *HSR: Health Services Research, 37*, 1403–1417.
- McCauley, R. J., & Swisher, L.** (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders, 49*, 34–42.
- McHorney, C. A., & Tarlov, A. R.** (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research, 4*, 293–307.
- Meline, T.** (2010). *A research primer for communication sciences and disorders*. Boston, MA: Pearson.
- Meline, T., & Paradiso, T.** (2003). Evidence-based practice in schools: Evaluating research and reducing barriers. *Language, Speech, and Hearing Services in Schools, 34*, 273–283.
- Meline, T., & Schmitt, J. F.** (1997). Case studies for evaluating statistical significance in group designs. *American Journal of Speech-Language Pathology, 6*(1), 33–41.
- Meline, T., & Wang, B.** (2004). Effect-size reporting practices in *AJSLP* and other ASHA journals, 1999–2003. *American Journal of Speech-Language Pathology, 13*, 202–207.
- Mintz, J.** (1972). What is success in psychotherapy? *Journal of Abnormal Psychology, 80*, 11–19.
- Morrison, D. E., & Henkel, R. E. (Eds.)**. (1970). *The significance test controversy*. Chicago, IL: Aldine.
- Murray, C. J. L.** (1994). Quantifying the burden of disease: The technical basis for disability adjusted life years. *Bulletin of the World Health Organization, 72*, 429–445.
- Nickerson, R. S.** (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods, 5*, 241–301.
- Nunnally, J. C., & Kotsche, W. E.** (1983). Studies of individual subjects: Logic and methods of analysis. *The British Journal of Clinical Psychology, 22*, 83–93.
- Ogles, B. M., Lunnan, K. M., & Bonesteel, K.** (2001). Clinical significance: History, application, and current practice. *Clinical Psychology Review, 21*, 421–446.
- Onslow, M., Jones, M., O'Brian, S., Menzies, R., & Packman, A.** (2008). Defining, identifying, and evaluating clinical trials of stuttering treatments: A tutorial for clinicians. *American Journal of Speech-Language Pathology, 17*, 401–415.
- Peterson, L. S.** (2008, February). "Clinical" significance: "Clinical" significance and "practical" significance are NOT the same things. Paper presented at the annual meeting of the South-west Educational Research Association, New Orleans, LA.
- Rosen, A., & Proctor, E. K.** (1981). Distinctions between treatment outcomes and their implications for treatment evaluation. *Journal of Consulting and Clinical Psychology, 49*, 418–425.
- Rosenbek, J. C., LaPointe, L. L., & Wertz, R. T.** (1989). *Aphasia: A clinical approach*. Boston, MA: College Hill Press.
- Schiavetti, N., Metz, D. E., & Orlikoff, R. F.** (2011). *Evaluating research in communicative disorders* (6th ed.). Boston, MA: Pearson.
- Schuele, C., & Justice, L.** (2006, August 15). The importance of effect sizes in the interpretation of research: Primer on research: Part 3. *The ASHA Leader*. Retrieved from www.asha.org/Publications/leader/2006/060815/f060815d/.
- Shriberg, L. D., & Lof, G. L.** (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics, 5*, 225–279.
- Shumak, I. C.** (1955). A speech rating sheet for stutterers. In W. Johnson & R. R. Leutenegger (Eds.), *Stuttering in children and adults: Thirty years of research at the University of Iowa* (pp. 341–347). Minneapolis: University of Minnesota Press.
- Sloan, J., Symonds, T., Vargas-Chanes, D., & Fridley, B.** (2003). Practical guidelines for assessing the clinical significance of health-related quality of life changes within clinical trials. *Drug Information Journal, 37*, 23–31.
- Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, R. B.** (2005). *Evidence-based medicine: How to practice and teach EBM* (3rd ed.). Edinburgh, United Kingdom: Elsevier Churchill Livingstone.
- Strupp, H. H., & Hadley, S. W.** (1977). A tripartite model of mental health and therapeutic outcome: With special reference to negative effects in psychotherapy. *American Psychologist, 32*, 187–196.
- Swanson, L. A., Fey, M. E., Mills, C. E., & Hood, L. S.** (2005). Use of narrative-based language intervention with children who have specific language impairment. *American Journal of Speech-Language Pathology, 14*, 131–143.
- Thompson, B.** (2002). "Statistical," "practical," and "clinical": How many kinds of significance do counselors need to consider? *Journal of Counseling & Development, 80*, 64–71.
- Thompson, M. S.** (1986). Willingness to pay and accept risks to cure chronic disease. *American Journal of Public Health, 76*, 392–396.
- Tingey, R. C., Lambert, M. J., Burlingame, G. M., & Hansen, N. B.** (1996). Clinically significant change: Practical indicators for evaluating psychotherapy outcome. *Psychotherapy Research, 6*(2), 144–153.
- Tukey, J. W.** (1991). The philosophy of multiple comparisons. *Statistical Science, 6*, 100–116.
- Wade, D. T., Smeets, R. J. E. M., & Verbunt, J. A.** (2010). Research in rehabilitation medicine: Methodological challenges. *Journal of Clinical Epidemiology, 63*, 699–704.
- Walker, S., & Rosser, R.** (1993). *Quality of life assessment: Key issues in the 1990s*. Boston, MA: Kluwer Academic.
- Walker-Batson, D., Curtis, S., Natarajan, R., Ford, J., Dronkers, N., Salmeron, E., ... Unwin, D. H.** (2001). A double-blind, placebo-controlled study of the use of amphetamine in the treatment of aphasia. *Stroke, 32*, 2093–2098.
- Wiklund, I.** (2004). Assessment of patient-reported outcomes in clinical trials: The example of health-related quality of life. *Fundamental & Clinical Pharmacology, 18*, 351–363.
- Wolf, M. M.** (1978). Social validity: The case of subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*, 203–214.

- 
- Woolf, G.** (1967). The assessment of stuttering as struggle, avoidance, and expectancy. *British Journal of Disorders of Communication*, 2, 158–171.
- World Health Organization.** (1980). *Classification of impairments, disabilities and handicaps*. Geneva, Switzerland: Author.
- World Health Organization.** (2001). *International classification of functioning, disability, and health*. Geneva, Switzerland: Author.
- Young, M. A.** (1976). Application of regression analysis concepts to retrospective research in speech pathology. *Journal of Speech and Hearing Research*, 19, 5–18.
- Young, M. A.** (1993). Supplementing tests of statistical significance: Variation accounted for. *Journal of Speech and Hearing Research*, 36, 644–656.

Copyright of American Journal of Speech-Language Pathology is the property of American Speech-Language-Hearing Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.