



Published in final edited form as:

J Commun Disord. 2006 ; 39(4): 266–291.

Single subject controlled experiments in aphasia: The science and the state of the science

Cynthia K. Thompson *

Aphasia and Neurolinguistics Research Laboratory, Department of Communication Sciences and Disorders, and Neurology, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, United States

Abstract

This paper discusses the use of single subject controlled experimental designs for investigating the effect of treatment for aphasia. A brief historical perspective is presented, followed by discussions of the advantages and disadvantages of single subject and group approaches, the basic requirements of single subject experimental research, and crucial considerations in design selection. In the final sections, results of reviews of published single subject controlled experiments are discussed, with emphasis on internal validity issues, the number of participants enrolled in published studies, operational specification of the dependent and independent variables, and reliability of measurement.

Learning outcomes—As a result of reading this paper, the participant will: (1) understand the mechanisms required for demonstration of internal and external validity using single subject controlled experimental designs, (2) become familiar with the basic requirements of single subject controlled experimental research, (3) understand the types of single subject controlled experimental designs that are the most appropriate for studying the effects of treatment for aphasia, and (4) become familiar with trends in the published aphasia treatment literature in which single subject controlled experimental designs have been used.

Studying the effects of aphasia treatment is an important scientific endeavor. As pointed out by Holland, From, DeRuyter, and Stein (1996) in a review of 200 papers addressing aphasia treatment in the mid-1990s, even though extant data show that aphasia treatment results in improved language, many questions concerning the efficacy of treatment for aphasia remain unanswered. Among the most important areas in need of research is determining the types of treatment for particular aphasic deficits that result in the greatest treatment effects.

As in all clinical domains, there are several approaches to studying the effects of treatment for aphasia. These include both group designs and single subject approaches. The purpose of this paper is to discuss the advantages of using single subject controlled experimental designs. A brief historical perspective is presented; advantages and disadvantages of single subject and group designs are discussed; basic requirements of single subject experimental research are presented; crucial considerations in design selection are discussed; and some of the problems inherent in published work in aphasia using single subject experimental design are addressed. This latter issue is discussed based on reviews of single subject controlled experiments by Kearns and Thompson (1991) and Thompson and Kearns (1991) as compared to subsequent reviews from 1989 to 1999 (Thompson, 2000) and from 2000 to 2005, undertaken for the purposes of this paper.

*Tel.: +1 847 491 2421; fax: +1 847 467 1036. E-mail address: ckthom@northwestern.edu.

1. A bit of history

Single subject experimental design and analysis has its roots in physiology and psychology, disciplines in which individual organisms served as the basis for discoveries including principles of learning, memory, sensation and perception, and the neural basis of language processing. Herman Ebbinghaus (1850–1909), using himself as a subject, was the first to investigate the properties of human memory. Using nonsense syllables as stimuli, he developed basic metrics of learning and forgetting, i.e., the learning curve, the forgetting curve, the serial position curve, and he invented several tests of memory, including measures of recall and recognition. Ivan Pavlov (1849–1936), interested in the physiology of the digestive system, also studied the behavior of individuals; i.e., studying the salivary response of individual animals to food when paired with other environmental stimuli, he discovered principles of conditioned learning. Gustav Fechner (1801–1887) developed measures for determining sensory thresholds and just noticeable differences (JNDs) in various sensory modalities by carefully studying responses of individual subjects to different stimulus intensities or locations. And in the domain of language, Paul Broca (1824–1880) study of one patient, Leborgne, in 1861 led to early ideas about language and brain function.

One important outgrowth of studying individual subjects was the discovery of individual differences, i.e., that not all humans respond the same way under particular stimulus conditions. Adolphe Quetelet (1796–1874), one of the most influential social statisticians from the nineteenth century, showed that human traits follow a normal curve, and that an average can be computed by studying a particular phenomenon in a group of individuals. Among other phenomena, he collected and analyzed data on crime and mortality, and influenced development of the Quetelet index, a measurement of obesity. His concept of the “average man”, as the ideal prototype of nature with deviations from it considered errors, was highly criticized. However, the fact that individual differences exist, that central tendencies can be computed, and that inter-subject variability (or error) can be accounted for or “averaged out” led, not surprisingly, to the use of mathematical methods to study human traits and served as the foundation for studying groups of individuals. Notably, it was Fechner who first applied statistical methods to psychological problems.

The work of Francis Galton (1822–1911) and Karl Pearson (1857–1936), which led to the notion of correlation, and later factor analysis, and Sir Ronald A. Fisher (1890–1962) who developed statistical procedures for comparing groups, including the analysis of variance and the split-plot technique, had considerable influence on the study of human behavior and the design of experiments. Interestingly, the procedures developed by Fisher grew from his work in genetics and agronomy. He conducted breeding experiments with mice, snails and poultry and studied the effects of various soil treatments, plant varieties, etc. on the yield of plots of plants at the Rothamsted Agricultural Experiment Station in the United Kingdom. In this work his interest was in gene dominance and, in his work with plants, variables that yielded the most fruitful crops. The fate of the individual organism was of little, or no concern.

1.1. Studying the effects of treatment for clinical problems

1.1.1. The case study method—The study of treatment effects for persons with disorders also has its roots in the study of individuals. Clinical patients suffering from behavioral or emotional disorders were noted to “recover” from these problems when provided with treatment. These observations led to clinicians' careful documentation of their procedures with individual patients. These case study reports implied that the treatment applied was what caused any notable clinical change in the patient. A classic case is that of Breuer and Freud (1957) who reported the “disappearance” of hysterical symptoms in Anna O. with hypnosis and “talking” therapy, which traced these symptoms to circumstances surrounding the death of the patient's father. As methods of scientific investigation began to emerge, such case reports were

strongly rejected for a number of reasons. First, these studies lacked experimental control, i.e., there could have been variables other than the treatment that influenced the changes noted, such as “spontaneous remission”, simply receiving attention from a clinician (i.e., the Hawthorne effect¹), or exposure to other events such as changes in family dynamics or participation in community activities, or any multitude of extraneous variables. Secondly, the methods applied were only loosely described and thus could not be replicated. Finally, the methods used to quantify change were often subjective, i.e., self-report or clinician judgment, and when tests or other measures were applied to measure change, the measures used were not valid, and the reliability of measurement was not considered.

Case studies also served as the starting point for investigating the effects of treatment for patients with aphasia. In an early case, Schuell and Sefer (1974) reported the results of a year of aphasia therapy for a gentleman, Mr. S., who presented with a “severe aphasia reflected in all language modalities” (p. 161), resulting from a cerebral hemorrhage. Extensive pretreatment testing revealed an aphasia marked by jargonistic output and poor auditory comprehension. He was able to tell his name and age, but could give no other biographical information; he used some short, high frequency, stereotypical social utterances and could produce some overlearned responses, such as counting to 10, but was unable to produce propositional sentences. He could name only one of 20 objects and could not repeat words or phrases. Auditory comprehension was severely compromised; he could follow only simple, high frequency commands, could only point to about 30% of objects named, and could point to no objects named in series of two or more. After one year of treatment, using “intensive stimulation of language” which employed a number of activities, he had made substantial improvements in language, noted as follows:

“[Following treatment] his speech was so fluent that most people would not have caught his few errors... [although] he admitted that he didn't always understand what was said to him... He scored in the 99th percentile on the Ammons Full Range Picture Vocabulary Test. Writing was still slow, but he was able to write good sentences... Reading was at the adult level, and [reading] comprehension was excellent (p. 167).”

While this and other case studies of aphasia treatment indicate improvement in language, they suffer from the same criticisms as those focused on treatment of psychological disorders. They were experimentally uncontrolled and, thus, one could not conclude with certainty that there was a functional relationship between the treatment provided and language improvement.

1.1.2. Alternatives to the case study approach—Because of the limitations of the case study approach as well as advances in statistical analysis, between-groups studies gained popularity. Early between groups studies of psychotherapy in the 1950s, however, showed no difference between experimental and control groups, leading to a rejection (by some) of this approach to treatment. However, after careful examination of the performance of individual subjects who participated in large outcome studies, Bergin (1966) reported that, in fact, some patients did improve with treatment, but others actually worsened. When subjected to statistical averaging of results, these differing performance patterns were averaged out, yielding an overall result of no treatment effect when compared to the control group. What these results demonstrated is that people, particularly those with disorders, are different from one another. Thus application of ill-defined, general treatment such as psychotherapy to a heterogeneous group of patients diagnosed with a particular disorder is not well suited for answering the more basic question on the effectiveness of a specific treatment for a specific individual, or even treatment for individuals who present with similar disorder profiles.

¹Improvement in a process resulting from observation of that process. The effect was first noted at the Hawthorne plant of Western Electric. Production increased not as a consequence of actual changes in working conditions introduced by the plant's management but because management demonstrated interest in such improvements.

In the area of aphasiology, between-groups experimental designs also became the scientific tool of choice in the 1960s–1980s. During that time at least 20 large-scale group treatment studies, enrolling 60 or more patients, and several smaller group studies including at least 10 participants, were published (see, for example, Basso, Capitani, & Vignolo, 1979; Vignolo, 1964; Wertz et al., 1986). Unlike the findings of psychotherapy studies, the results of these group studies largely showed that treatment results in general improvement in language, although, in some cases the treatment effects were not particularly strong or were lacking (Lincoln et al., 1984). However, like the psychotherapy literature, the treatment provided was only vaguely described, for example, Vignolo (1964) provided “traditional” treatment, or was unspecified (Lincoln et al., 1984), and often the clinicians delivering the treatment varied within studies. Therefore, it is impossible to glean from such studies what treatment variables affected language improvement, and, in addition, the treatment is impossible to replicate. Further, many published group treatment studies failed to include a control group, thus spontaneous recovery and other variables may have influenced the language improvement noted. Finally, patients with different language profiles and differing levels of severity were entered into treatment. Thus, it is likely that not all patients improved to the same extent. While this is not a problem, per se, for group studies since the primary means for determining treatment effects is to compare experimental to control groups, and variability inherent in both groups is managed through statistical methods, the issue is that such studies tell us nothing about which subjects performed better or worse than others, and why.

1.1.3. Single subject experimental research—Given the limitations inherent in group studies, researchers in clinical psychology, education and other disciplines began using single subject experimental designs to study the effects of treatment. These designs, like case studies, afford careful analysis of treatment effects in individual subjects. However, unlike the case study approach, single subject experimental designs provide a mechanism for controlling the influence of extraneous variables, or for demonstration of internal validity—to isolate the independent variable (treatment) as responsible for the experimental effects. In this latter regard, single subject designs are like between-groups designs. However, rather than comparing performance between experimental and control groups of subjects as in between-groups designs, the strategy in single subject experimentation is to compare performance between experimental and control conditions (or phases) in individual subjects. Another difference between case studies and single subject controlled experiments is that single subject designs require replication of treatment effects both within and across subjects, thus they are *not* synonymous with $N = 1$ experiments. This myth that single subject experimental analysis requires studying a “single subject” has perhaps retarded use of these designs in some disciplines.

It was not until the early 1980s that this method was used to examine the effects of treatment for aphasia and other communication disorders. The first published paper in aphasia using a single subject experimental design was that by Thompson and Kearns (1981) (see Thompson, Kearns, & Edmonds, in press). The first book detailing this methodology for studying communication disorders was published in 1983 (McReynolds & Kearns, 1983), and the first series of papers to appear in a journal concerned with communication disorders was that by McReynolds and colleagues in the *Journal of Speech and Language Research* (Connell & Thompson, 1986; Kearns, 1986; McReynolds & Thompson, 1986).

2. Selection of an experimental approach for studying the effects of treatment

There are several issues relative to selecting an appropriate experimental approach, i.e., group versus single subject, to study the effects of aphasia treatment. These include inter- and intra-subject variability and averaging of results, the generalizability of results (or external validity),

generalization of treatment effects across responses and stimulus conditions, ethical issues, and practical problems. These issues are addressed below.

2.1. Variability and averaging results

One issue, as discussed above, pertains to the averaging of results, which essentially obscures individual clinical outcomes. It is well known that individuals with aphasia are heterogeneous, i.e., inter-subject variability is a given. Even those diagnosed with a certain type of aphasia, such as Broca's aphasia or Wernicke's aphasia, present with varying language deficit patterns. And even patients carefully selected for their language deficit profile are seldom, if ever, truly homogeneous. They can, and often do, vary markedly with regard to neurological as well as psychosocial and other variables. Given this heterogeneity, it is likely that treatment effects differ across individuals. Thus, averaging of results, which is required in group experimental studies, may be contraindicated, particularly in early stages of experimentation when the effects of a particular treatment are unknown. Once treatment effects have been established in individual subjects, group designs, examining the effects of treatment in larger numbers of participants are more appropriate.

Individuals with aphasia also often show intra-subject variability, i.e., performance varies from day to day. One advantage of the single subject experimental approach is that intra-subject variability is transparent as the experiment unfolds because repeated measurement of the dependent variables across phases of the study is required (see below for further discussion of repeated measurement). Alternatively, in the group approach, such variability is hidden. Performance on the dependent variables is usually measured only twice—once before treatment and one following treatment application.

2.2. Generality of results

An important consideration in treatment research is the generality, or external validity, of findings. On the face of it, group designs, rather than single subject designs, would appear to be the method of choice if the goal is to “generalize” the results of a study to the population. Inferential statistics can elegantly be used to estimate such generality. However, there are problems here, particularly with regard to clinical research. One of the many assumptions underlying inferential statistical tests is that participants included in the study sample be randomly selected from the population. Thus, it is only when participants are randomly selected (and randomly assigned to experimental versus control groups) that statements generalizing the results from the sample to the population can be made. In aphasia research it is extremely difficult, if not impossible, to adequately sample the population. This, of course, would require that all individuals with a certain aphasia profile be equally likely to be selected and that they would be available to participate in the study.

In most group research in aphasia, study participants are recruited from a particular clinical facility or facilities, or from a particular geographical area, to meet certain inclusionary criteria. In such studies, while hypothesis testing is still possible using statistical methods, the significance statements are restricted to the effect of the experimental treatment on the study participants; generalization to other aphasic individuals, even those who present with language deficit patterns similar to those who served as study participants cannot be made based on statistics. Generality statements can only be made based on logical non-statistical considerations.

Another problem with inferential statistics, even if the study sample is randomly selected from the population, is that statements of generality can only be made to the population from which the sample was selected. These methods do not afford statements of generality to individual patients. In addition, the assumption underlying true random selection is that all possible

variations of the study population will be represented in the study sample, thereby, diminishing the relevance of the group effect to a specific individual.

Single subject experimental studies often are criticized because they are thought to not yield externally valid results, i.e., the results cannot be generalized beyond the study participants. This statement is true if one relies on inferential statistics to estimate the generalizability of findings. Indeed, such statistical applications are inappropriate for use with data derived from single subject experimentation for a number of reasons including serial dependency (see Kazdin, 1982; Gorman & Allison, 1996, for discussion of the use of statistics with single subject data). However, there are methods, albeit non-statistical ones, for estimating external validity in single subject experimental research. These include direct and systematic replication. As mentioned above, any properly executed single subject experiment requires direct replication of the treatment effect both within and across participants. For example, in a successful A–B–A–B experiment, the treatment effect is replicated two times (during the B phase) in each participant in the study. Thus, if four individuals participated in the study with successful outcome, this would equal eight replications of the treatment effect. The experimental effect is then established in additional experiments with homogeneous participants, further increasing the number of replications of the effect. Several direct replications then establish the generality of findings to similar patients. The logic is simple: the greater the number of replications, the greater the generality of the effect. This logic is no different than logical, non-statistical generality derived from between-groups studies of homogeneous subject, in which random selection and assignment to experimental and control groups is absent.

The other method for estimating generality of findings from single subject studies is systematic replication. For example, once a treatment effect has been adequately established through direct replication, the next step is to systematically examine the effects of the treatment with patients who differ from the original participants in language deficit patterns, severity, or in other ways, and/or under different conditions, e.g., clinician delivered versus computer delivered conditions. The goal of systematic replication is in essence the same as that in powerful factorial designs, undertaken with groups of participants. Such designs not only establish the effects of treatment, but also evaluate what type of patients, under what conditions, improve with a given treatment.

2.3. Investigation of generalization patterns

Generalization is without doubt one of the most important outcomes of successful treatment for aphasia. In fact, without evidence of generalization, treatment may be deemed ineffective. Developing treatments for aphasia that result in maximal generalization has been a goal of aphasia treatment researchers for several decades and recently some principles for promoting generalization have been discovered. Notably, these principles have resulted from careful study of patterns of acquisition and generalization using single subject experimental designs. Because repeated measurement is a requirement of these designs, they are well suited to examine generalization as it emerges during treatment. This is not to say that generalization cannot be examined using a group design strategy, because it can, of course. However, as pointed out above, group designs generally measure the dependent variables under study (including those designed to examine generalization) once prior to treatment and once following treatment. Thus, generalization, as it unfolds throughout treatment cannot be analyzed.

There are two types of generalization—both of which can be included as dependent measures in single subject controlled experiments with aphasic patients. The first is response generalization, i.e., changes in untrained language behavior. For example, when treatment focused on improving retrieval of a set of words results in improvement in the ability to retrieve an untrained set of words, response generalization has occurred. Response generalization can be tested to linguistically related and/or functionally related responses.

The second type of generalization is stimulus generalization, i.e., changes in untrained language conditions. This results when, for example, treatment focused on improving word retrieval impacts the ability to retrieve words in narrative discourse conditions, or to use them for functional communication in naturalistic settings. Both types of generalization can, and should, be systematically measured using single subject experimentation.

2.4. Ethical concerns

Another issue relevant to group versus single subject experimentation is that the group approach requires a control group—generally a no-treatment control group. Of course, the ethical concern here is withholding treatment from patients. While, in theory, withholding an experimental treatment, the effects of which are unknown, from a patient is probably not unethical, the idea of withholding treatment, even inadequate treatment, is not popular among clinicians, aphasic individuals and their family members. There are ways, however, in group research to circumvent this problem, i.e., a cross over design can be used whereby treatment is withheld for the control group only until the experimental group has completed treatment. Then treatment is provided for the control group. However, even though this and other solutions to the problem are possible, there are many group treatment studies in aphasia in which control groups are lacking, or control groups are comprised of patients who are unable to participate in treatment for reasons such as motivation, family support, transportation, etc. Both of these situations are problematic from a scientific point of view. Clearly the control group issue is overcome by using a single subject experimental approach, in which control groups are not required by the design.

2.5. Practical problems

Recruiting individuals with aphasia to participate in research is no easy task. As discussed above, random selection and assignment of aphasic patients to experimental and control groups often is unrealistic. In addition, patients with aphasia do not comprise a homogeneous population. Thus, obtaining sufficient numbers of individuals who are matched on relevant variables to constitute the required groups in group research is often not possible. Studying large groups of patients can also be expensive and very time consuming, particular in the arena of treatment research. Treatment research in aphasia requires delivery of treatment often for several weeks or even months and analysis of such data is time consuming. Indeed, single subject experimental research also is expensive and time consuming; however, generally fewer participants are required as compared to group approaches.

3. Requirements of single subject experiments

The basic requirements of single subject experiments investigating the effects of treatment for aphasia are identical to those for group experiments. As discussed by Chambless and Hollon (1988), these include demonstration of experimental control or internal validity, operational definition of variables, repeated measurement, reliability of measurement, and clear description of study participants.

3.1. Establishing internal validity

The basic processes involved in developing internally valid single subject controlled experiments involves development and arrangement of design phases such that the effects of extraneous variables on outcome measures can be ruled out. Depending on the particular single subject design selected, phases also must be arranged to control for order effects and other threats to internal validity. In addition, as mentioned above, design phases need to be arranged such that replication of effects within and across participants can be accomplished.

So what are these design phases? Design phases are defined as periods of time during which a particular condition of the experiment is in effect. These include both control (baseline) and experimental (treatment) phases. During both phase types, the dependent variable(s) of interest are measured, but only during the experimental phases is the independent variable applied. Comparing performance across phases of the study reveals the treatment effect, e.g., the behaviors under study are low and stable during control phases, and change is in a positive direction during experimental phases. There are a number of elegant ways in which these phases can be arranged, which depends on a number of issues, including the behavior under study. Phase arrangements, or types of design, that are most appropriate for studying the effects of treatment for aphasia are discussed below.

3.2. Definition of variables

All experimental studies require operational specificity of the independent (treatment) and dependent (outcome) measures. Operational definition of the independent variable in treatment research involves a detailed description of the treatment that is under investigation. What visual or auditory stimuli will be used, how are the stimuli delivered, what are the response parameters, including response time, what feedback is provided, and so on? In order for the treatment to be replicated, for both research and clinical purposes, it must be precisely described. The dependent measures also require precise description. How is the outcome of treatment to be measured? A common practice in single subject research is to develop daily probe tasks, measuring the language behavior under study in conditions where no feedback is provided. These can include both on-line measures such as reaction time and off-line measures, depending on the goal of a particular experiment. It is the clients' responding to these probe tasks that serve as the dependent variable throughout the study.

3.3. Repeated measurement

What is required in single subject experiments, which is not in group designs, is repeated measurement of the dependent variable(s) throughout the course of the experiment. (Group studies require that dependent measures be repeated only once—following the completion of treatment, with performance from pre-treatment and post-treatment testing compared to one another.) This is accomplished by administration of the dependent measures regularly and using identical procedures throughout both control (baseline) and experimental phases of the experiment. This allows close inspection of behavioral variability, as well as the slope, level and trend of the dependent variables across phases of the study.

3.4. Reliability of measurement

One other important aspect of treatment research concerns the reliability of measurement of both the independent and dependent variables included in the experiment. Treatment research largely involves observation of human behavior; inherent in human observation is human error, as well as observer bias. Although it is difficult, if not impossible, to overcome human error, estimates of reliability can be made using measures of inter-observer agreement. Such measures involve use of an independent observer who, together with the primary experimenter, score important events in the study, including details pertaining to delivery of treatment (reliability on the independent variable) and responses made on the dependent measures (reliability on the dependent variable). With regard to the independent variable, the observer quantifies salient aspects of treatment, e.g., counts of the number of experimental trials delivered per treatment session, scores of adherence to procedural detail within trials, etc. Reliability on the dependent variable involves scoring of participant responses on the probe task(s) based on established criteria. When the independent observers agree to a high degree, it is unlikely that human error or observer bias is operating, adding an element of believability to the data. Lack of agreement between observers alerts the experimenter to problems with the

experiment, e.g., imprecise operational definitions of the study variables, which if discovered early in the course of an experiment can be modified.

3.5. Description of study participants

Careful selection of participants based on explicit inclusionary and exclusionary criteria is a requirement in all research using human subjects. This is important for internal and external validity purposes and it also impacts the extent to which an experiment can be replicated by other investigators. Without proper control of subject variables, the internal validity of the experiment suffers because it is possible that observed effects resulted not only from manipulation of the independent variable(s), but also from uncontrolled subject variables. Participant selection is also relevant to external validity, i.e., as noted above, generalization of results either using statistical applications or non-statistical logical approaches depends on the characteristics of the study participants. Finally, the results of any study of human behavior should be replicable, both in future scientific studies and in the clinical domain. Without adequate description of subjects, such replication would be compromised. Brookshire (1983) suggested that in aphasia research, certain basic information about study participants be provided including age, education, handedness, source of subjects, etiology, lesion location, time post stroke, severity of aphasia, and type of aphasia. In addition to these data, it is suggested that a careful description of the language and other cognitive abilities of patients entered into the study be provided since individuals labeled with a certain type of aphasia often show different deficit patterns.

4. Developing an appropriate design for studying the effects of aphasia treatment

As discussed above, developing a single subject experimental design involves arrangement of control and experimental phases, such that the effects of the treatment variable, and not extraneous variables, can be determined, and such that replication of the treatment effect can be established. Usually, control phases are labeled A phases, while experimental phases are labeled B.

4.1. Arranging phases

A traditional design is the A–B–A–B design. In this design, the behavior(s) under study are first measured in the baseline (A phase), then treatment is applied in the first B phase, the treatment is then withdrawn in the second A phase, and finally treatment is reapplied in the second B phase. In order to demonstrate experimental control using such a design (a) performance on the dependent variable must be stable and low (or high depending on the behavior under study) during the first A phase, (b) a change in the dependent variable(s) must be seen when comparing performance in the first A phase with that in the first B phase, (c) the dependent variable(s) must reverse in the second A phase, i.e., return to baseline levels, and (d) during the second B phase, the treatment effect must be re-established, i.e., change in the dependent variable(s) is once again seen. This sequence of events allows within subject replication and when shown in several participants, across subject replication is established.

The requirement that the dependent variable(s) return to baseline levels in the second A phase presents a potential problem for treatment research in aphasia, since the goal of treatment research generally is discover methods to improve language function. If the treatment is successful reversal is not desirable nor may it be possible, i.e., a simple withdrawal of treatment should not result in unlearning. While there are methods for “forcing a reversal”, i.e., training error responses during the second A phase, this practice in aphasia treatment is not recommended.

The A–B–A–B design, as pointed out by Kearns (1986), is better suited for studying behaviors that are likely to, or are expected, reverse following the withdrawal of treatment. For example, Renvall, Laine, and Martin (in press) used an A–B–A design to examine the effects of contextual priming in a patient with anomia. Their hypothesis was that contextual priming would improve naming when applied, but that when withdrawn, no long-term effects of treatment would be noted. As shown in Fig. 1, this is exactly what they found. A return toward baseline levels of performance occurred when treatment was withdrawn.

A frequently used alternative to the A–B–A–B design is the multiple baseline design across behaviors, which does not require returns to baseline levels of responding to demonstrate internal validity. This design, in essence, is a series of A–B designs. Baseline data are collected on two or more independent behaviors in each study participant. Following this, the independent variable is applied to one behavior at a time during the B phase, while the A phase is continued for the remaining behaviors under study. When a treatment effect is established for the first behavior, treatment is extended to the second behavior, and so on, until all behaviors have been individually treated. Experimental control is demonstrated in this design when changes in the dependent variable(s) occur only when the B phase is in effect for each behavior; baseline performance of untreated behaviors remains stable, until treated. Like the A–B–A–B design, the multiple baseline strategy allows replication of the treatment effect within subjects, i.e., each application of treatment to a different behavior instantiates a replication. Across subject replication is established by entering more than one participant into the study.

Because the multiple baseline design requires sequential application of treatment to separate behaviors order effects must be ruled out. Thus, application of treatment to the behaviors under investigation must be counterbalanced across participants and the number of participants required for a particular study is dependent on the number of behaviors studied. Take for example, a study examining the effects of treatment on naming. The experimenter decides to study three sets of words, with each set tested in baseline and sequentially trained. To rule out order effects in such a study, the order of training each word set must be counterbalanced; thus three participants are required, with each receiving a different treatment order. For full replication in such a study, each order requires replication in an additional participant, thus a multiple baseline design across behaviors with three behaviors requires six participants.

Another issue relevant to the multiple baseline design across behaviors is that the behaviors must be functionally independent and, at the same time, amenable to the treatment under investigation. This means, for example in a naming study using three sets of words, that training one set of words would have no effect on the untreated sets. If the behaviors are not functionally independent, treatment of one set may influence the others, i.e., generalization may occur across sets. While such an effect is often desired as noted above, it is a disaster experimentally, i.e., experimental control is lost. If it is unknown whether or not the behaviors under study are functionally independent, or if one of the goals of the research is to examine generalization for example across word sets, then a multiple baseline across participants can be combined with the multiple baseline across behaviors. In this case, treatment is sequentially applied across participants following baselines phases of increasing length. The logic here is that treatment will affect behavior when and only when it is applied. Thus, if it is the treatment, and not extraneous variables, that are responsible for the behavioral effect, no change will be seen for any participant during baseline, regardless of its length. This extra design component serves as an insurance policy; if generalization occurs, experimental control is maintained.

Fig. 2 shows the results of treatment for one (of four) participants in a study using a multiple baseline design strategy by Kiran and Thompson (2003). The purpose of the experiment was to examine the effects of semantic complexity (as determined by the typicality of items entered into treatment) on naming in fluent aphasia. Three sets of words were selected from two

semantic categories. The word sets were comprised of either typical items within the category (e.g., *robin* is a typical bird), items of intermediate typicality, and atypical items (e.g., *ostrich* is an atypical bird). The hypothesis of the study was that training less semantically complex items (i.e., typical items) within a category would have no effect on more complex items (i.e., intermediate and atypical items), but that when atypical items were trained, generalization to items of lesser complexity (i.e., intermediate and typical items) would be forthcoming. Using a multiple baseline design across behaviors, one set of words was trained at a time while generalization was observed to the untrained sets. When generalization did not occur, treatment was extended to the next set, and so on, until all sets were trained. Fig. 2 shows no generalization across behaviors when typical items from the first category (birds) were trained; thus all sets were sequentially entered into treatment with a treatment effect noted for each set. For the second category (vegetables) treatment was first applied to atypical items and generalization to untrained items was noted. This same pattern was noted across participants. Thus experimental control was accomplished when the order of treatment was typical–intermediate–atypical, but was lost when the order was switched to atypical–intermediate–typical. Since this was an expected outcome of the study, an additional design component was included, i.e., a multiple baseline across participants. Therefore, even though generalization occurred with the latter treatment schedule, experimental control was demonstrated.

There are other types of single subject controlled experimental designs that can be used to examine the effects of treatment, however, the multiple baseline strategy is the most commonly used and is likely the best suited for most studies of aphasia. One important point is that studying the differential effects of two (or more) treatments is difficult within a single subject experimental framework. While this can be done using an Alternating Treatments Design (ATD), it is difficult to control for carry over effects from one treatment to another. In such a design, each study participant simultaneously receives all treatments under investigation in an alternating manner, with the order of presentation of each treatment counterbalanced across treatment sessions and participants, e.g., all treatments are applied consecutively during one treatment session, with the order changed in subsequent sessions. Following daily delivery of each treatment, the dependent measures are administered. For example, consider a study examining the effects of two different approaches for improving word retrieval. Each day, both treatments are applied and word retrieval is measured following each. The problem is that, given that word retrieval improves, there is no way to know which treatment (or whether both) caused the improvement. While there are ways to overcome this problem, for example, each treatment can be applied to functionally independent language behaviors (e.g., sets of linguistically unrelated words), carry over effects from one treatment to another is impossible to control.

Fig. 3 shows the results of a study using an ATD, combined with a multiple baseline design across behaviors, reported by Wambaugh (2003). The purpose of the study was to examine the differential effects of two lexically based treatment approaches to naming, a phonological cueing treatment (PCT) and a semantic cueing treatment (SCT), in a patient with chronic anomia. Following baseline, the two treatments were simultaneously applied to two separate word sets (sets 1 and 2), while sets 3 and 4 remained in the baseline condition. Once an acquisition effect was noted for the first two sets, the two treatments were simultaneously applied to the second two word sets. As can be seen experimental control was demonstrated, since improved naming was noted for all trained sets when treatment was given, but not during baseline testing. What is unclear is which treatment resulted in the treatment effects. It is possible that both treatments were effective, with the semantic treatment showing a slightly stronger effect, as concluded by Wambaugh (2003). Or it is possible that only one of the treatments was effective and that generalization occurred from the effective treatment to the word set assigned to the other treatment. While this is unlikely since there was no generalization effect of treatment from sets 1 and 2 to 3 and 4, it a potential explanation. The point here is

that it is not possible to completely rule out carry over effects using an ATD. Thus, in general, group designs are better suited for examining the differential effects of treatment.

5. The state of single subject controlled experimentation in aphasia treatment research

In this section, published reports of aphasia treatment in which single subject controlled designs were used are considered. Papers examining the effects of aphasia treatment published in the *American Journal of Speech and Language Pathology* (AJSLP), *Aphasiology*, *Brain and Language*, and the *Journal of Speech, Language, and Hearing Research* (JSLHR) from 2000 to 2005 were reviewed and sorted by design type: case study, single subject experimental, or group design. The papers using single subject controlled designs were further analyzed for their technical merit and compared with findings derived from previous reviews of single subject research in aphasia treatment by Kearns and Thompson (1991) and Thompson and Kearns (1991) from 1978 to 1987², and by Thompson (2000) from 1988 to 1999³. Several aspects of single subject controlled designs were evaluated, including (a) demonstration of experimental control or internal validity, (b) replication within and across participants, (c) operational specificity of independent and dependent measures, and (d) reliability of measurement.

5.1. General trends in treatment research in aphasia from 2000 to 2005

As can be seen in Fig. 4 a total of 81 studies examining the effects of aphasia treatment were published between the years 2000 and 2005 (see Appendix B for a listing of studies reviewed). Of these studies, 30 (37%) were case studies; 40 (49%) were single subject controlled experiments, and 11 (14%) were group experimental studies. Trends over the years show that in most all years there were greater numbers of single subject experiments than the other design types; a high number of case studies with no experimental control or internal validity were found; and fewer group studies were found as compared to the other design types. With respect to the frequency of single subject controlled and group experiments, these data are in keeping with the trends reported in the 1978–1987 review by Kearns and Thompson (1991). Single subject experiments increased from 17% in 1978 to 80% in 1987 with an average of 40% across the years under review, whereas, group studies were far fewer, between 0 and 17% over the years (see p. 35).

The situation regarding case studies, however, is surprising. Kearns and Thompson (1991) found a marked decrease in case studies with 75% in 1978 and 20% of all treatment studies in 1987 being case studies. Kearns and Thompson were encouraged by this trend, suggesting that it could be “viewed as evidence that investigators are opting for a more powerful design technology for conducting treatment studies” (p. 37). That case studies remain prevalent in the literature, however, suggests that both researchers and journal editorial boards remain unconvinced that powerful designs affording experimental control are required in aphasia treatment research. This is not to say that case studies do not have their place in the aphasia treatment literature. They are particularly worthy of publication if a novel approach to treatment is discussed, which can later be tested in a controlled study. However, the case studies in the aphasia literature are largely not of this type.

Notably, when examined by the journal in which aphasia treatment research has been published, the majority of papers appeared in *Aphasiology*, including 25 case studies, 32 single subject controlled studies, and 10 group studies. Far fewer papers were found in *AJSLP*, *Brain*

²This review was limited to papers appearing in the Proceedings of the Clinical Aphasiology Conference.

³This review included papers published in *AJSLP*, *Aphasiology*, *Brain and Language*, and *JSLHR*.

and Language, and *JSLHR*. Notably, no group or case studies were found in *AJSLP* or *JSLHR*; only single subject controlled experiments were found in these journals (see Fig. 5).

5.2. Experimental control

In review of the single subject controlled aphasia treatment studies, Fig. 6 shows that 62% of studies published between 1878 and 1987 demonstrated experimental control. This percentage decreased somewhat to 52% from 1988 to 1999, but increased to 75% in the latest review from 2000 to 2005. While the trend toward well-controlled experiments over the years is encouraging, these data indicate that the mere use of a single subject design does not constitute an experimentally controlled study. One common problem is inherent in the use of A–B–A–(B) designs. As noted above, such designs require a return to baseline level performance in the second A phase; when this does not occur, the design is in essence an A–B with maintenance testing in the second A phase. In this case, the influence of extraneous variables on patient performance cannot be ruled out. For example, change from the A to B phase could have resulted from spontaneous recovery of from a myriad of other variables that occurred at the same time that treatment was applied. Several researchers have encountered this problem (e.g., Bernstein-Ellis, Wertz, & Shubitowski, 1987; Boyle & Coelho, 1995; Le Dorze, Jacob, & Coderre, 1991), thus, once again, it is suggested that these designs be avoided in aphasia treatment research (at least in isolation, without additional design components introduced for experimental control purposes).

5.3. Replication within and across participants

These reviews also found that not all studies include enough participants to establish replication within and across participants. As noted above, single subject controlled experimentation does not mean that studying only a single subject is required. Fig. 7 shows that in most early studies (1978–1987) only one subject was studied (78%), and while this trend decreased from 1988 to 1999, the most recent review showed that many researchers continue to fail to include sufficient numbers of participants for replication purposes. In the 2000–2005 review, 38% of published studies included only one participant. Fig. 7 also shows that while some studies have included more than one participant only 18%, 28% and 35% of studies across reviews, respectively, included appropriate numbers of participants to demonstrate proper replication. As noted above, such replication is essential for both internal and external validity purposes—replication within participants addresses internal validity, whereas replication across participants addresses external validity. As the number of replications across participants increases, confidence that the effects of the study can be generalized also increases.

5.4. Operational definition of variables

Operational definition of variables was evaluated in the 1988–1999 and 2000–2005 papers only. In these two series, the extent to which the dependent and independent variables were described such that they could be replicated either for research or clinical purposes indicated a general trend toward better definition of these variables. Most studies (89%) included explicit description of probe tasks used to evaluate the effects of treatment, and most included repeated measurement across all phases of the study. The independent variable was described better in some studies than in others, but virtually none of the studies used terms such as “traditional treatment” and all reports included at least rudimentary information about the stimuli used in treatment as well as the treatment protocol.

5.5. Reliability of measurement

Over the years an increase in the number of studies including reliability data on the dependent variable also was noted. In the 1978–1987 review only 22% of studies reported such data as compared to the 1988–1999 and 2000–2005 review in which 58% and 64% of studies,

respectively, reported reliability on the dependent measures (see Fig. 8). While this upward trend is encouraging, these data show that not all single subject experimental researchers, even in the most recent review, are gathering reliability data as part of their experiments.

The data concerning reliability on the independent variable are disappointing. As can be seen in Fig. 8, none of the studies from 1978 to 1987 reported such data, and from 1988 to 1999 only 15% reported such data, and from 2000 to 2005 only 18% of studies included reliability on the independent variable. As noted above, these data provide a measure of assurance that the treatment was applied as described. Indeed, adherence to experimental protocols in treatment research is essential such that the treatment effect can be attributed to its application as described by the researcher.

6. Summary and conclusion

This paper has emphasized the rationale behind single subject controlled experimentation, provided some critical distinctions between group and single subject approaches, detailed the technical requirements of single subject designs, and presented trends in the aphasia treatment literature, where single subject controlled designs have been utilized.

While single subject controlled experiments are often used in aphasia treatment research, there remain a larger number of case studies, which lack the experimental control that permits valid conclusions to be drawn about the functional relationship between independent and dependent variables. In addition, many published single subject controlled experiments have been unsuccessful in establishing experimental control, particularly those in which an A–B–A–(B) design is employed. Further, some researchers continue to include only one participant in their studies and even when more than one participant is studied, the total number is inadequate for proper within and across subject replication. Finally, it was noted that while operational definitions of dependent and independent variables often are reported in published studies, not all studies include reliability data on the dependent measures, and very few include reliability on the independent variable. Future research using single subject controlled designs could benefit from consideration of these requirements.

In spite of the shortcomings found in some studies, numbers of well-designed experiments are available which attest to the applicability of single subject controlled designs to the study of aphasia treatment. Using these designs, carefully selecting study participants for the specific nature of their language impairment, precisely describing the components of treatment as well as the outcome measures and carefully gathering reliability data, researchers have discovered treatments that are effective for patients with certain types of language impairments. As pointed out by Holland et al. (1996), the field of aphasia treatment research is well beyond asking and answering the question of general treatment efficacy, in that we have known for many years that treatment does improve language ability. What we need to know now is what treatment (s) to provide for what disorder(s). Progress toward answering this question has clearly been made.

Many studies also have addressed the generalization of treatment effects and have shown that treatment often impacts both untrained language behaviors (response generalization) and language use in untrained conditions (stimulus generalization). Indeed, knowing the effects of specific types of treatment for patients with certain language impairments as well as understanding the extent to which these treatments result in generalized language use is important particularly in the current health care climate, which imposes limitations on the treatment that can be provided.

This is not to say that our work is done, as further work detailing the effects of certain treatments for certain language deficits is needed. Direct and systematic replication of established effects

also is necessary in order to enhance the generalizability of findings. In addition, there are many other unanswered questions regarding aphasia treatment effects. Among these are: When is treatment most effective? How often should treatment be provided? What treatment(s) work best for what language deficit(s)? The answers to these and other questions await further research, some of which can be elegantly addressed using single subject controlled experimental designs.

Appendix A. Continuing education

Study Questions and Answers (correct answers are bolded)

1. Case studies and single subject controlled experiments:
 - a. Are similar in that both require studying only one participant.
 - b. Are different in that internal validity is not demonstrated in case studies, but is a requirement of well designed and executed single subject experiments.
 - c. Are appropriate design choices for experimentally investigating the effects of treatment for aphasia.
 - d. Are both inferior to group experimental research designs.
2. Group experimental and single subject controlled experimental designs are similar to each other in the following ways:
 - a. Both entail mechanisms for controlling for the effect of extraneous variables on the dependent measures.
 - b. Both require operational specificity of the independent and dependent variables.
 - c. Both require measurements of reliability on the independent and dependent variables.
 - d. All of the above.
3. Multiple baseline designs across behaviors:
 - a. Are internally valid even if generalization across behaviors occurs.
 - b. Cannot be used to examine generalization across behaviors even when combined with other single subject experimental designs.
 - c. Can be used to examine generalization across behaviors when combined with other single subject experimental designs.
 - d. None of the above.
4. Which of the following can potentially affect the internal validity of a single subject controlled experiment?
 - a. Inadequate selection criteria of study participants.
 - b. Failure to demonstrate a reversal of the treatment effect in an A–B–A–(B) design.
 - c. Generalization across behaviors in a multiple baseline design across behaviors design.
 - d. All of the above.
5. External validity using single subject controlled experimental designs:

- a. Is addressed through direct and systematic replication of treatment effects.
- b. Cannot be addressed because study participants are not randomly selected.
- c. Is addressed using parametric statistics as in group designs.
- d. Is considered unimportant since the goal is to establish the effects of treatment in a single subject.

Answer key:

- 1. (b)
- 2. (d)
- 3. (c)
- 4. (d)
- 5. (a)

Appendix B

Published single subject controlled experiments from 2000 to 2005.

- Beeson PM, Hirsch FM, Rewega MA. Successful single-word writing treatment: Experimental analyses of four cases. *Aphasiology* 2002;16:473–491.
- Boucher V, Garcia LJ, Fleurant J, Paradis J. Variable efficacy of rhythm and tone in melody-based interventions: Implications for the assumption of a right-hemisphere facilitation in non-fluent aphasia. *Aphasiology* 2001;15:131–149.
- Boyle M. Semantic feature analysis treatment for anomia in two fluent aphasia syndromes. *American Journal of Speech-Language Pathology* 2004;13:236–249. [PubMed: 15339233]
- Clausen NS, Beeson PM. Conversational use of writing in severe aphasia: A group treatment approach. *Aphasiology* 2003;17:625–644.
- Coelho CA. Direct attention training as a treatment for reading impairment in mild aphasia. *Aphasiology* 2005;19:275–283.
- Coelho CA, McHugh, Boyle M. Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology* 2000;14:133–142.
- Conley A, Coelho CA. Treatment of word retrieval impairment in chronic Broca's aphasia. *Aphasiology* 2003;17:203–211.
- DeDe G, Parris D, Waters G. Teaching self-cues: A treatment approach for verbal naming. *Aphasiology* 2003;17:465–480.
- Fink RB, Brecher A, Schwartz MF, Robey RR. A computer-implemented protocol for treatment of naming disorders: Evaluation of clinician-guided and partially self-guided instruction. *Aphasiology* 2002;16:1061–1086.
- Fox LE, Sohlberg MM, Fried-Oken M. Effects of conversational topic choice on outcomes of augmentative communication intervention for adults with aphasia. *Aphasiology* 2001;15:171–200.
- Freed D, Celery K, Marshall RC. Effectiveness of personalised and phonological cueing on long-term naming performance by aphasic subjects: A clinical investigation. *Aphasiology* 2004;18:743–757.
- Fridiksson J, Holland AL, Beeson P, Morrow L. Spaced retrieval treatment of anomia. *Aphasiology* 2005;19:99–109. [PubMed: 16823467]
- Gold M, VanDam D, Silliman ER. An open-label trial of bromocriptine in nonfluent aphasia: A qualitative analysis of word storage and retrieval. *Brain and Language* 2000;74:141–156. [PubMed: 10950911]
- Hickey EM, Bourgeois MS, Olswang LB. Effects of training volunteers to converse with nursing home residents with aphasia. *Aphasiology* 2004;18:625–637.
- Hopper T, Holland A, Rewega M. Conversation coaching: Treatment outcomes and future directions. *Aphasiology* 2002;16:745–761.

- Jacobs BJ, Thompson CK. Cross-modal generalization effects of training noncanonical sentence comprehension and production in agrammatic aphasia. *Journal of Speech, Language, and Hearing Research* 2000;43:5–20.
- Kiran S. Training phoneme to grapheme conversion for patients with written and oral production deficits: A model based approach. *Aphasiology* 2005;19:53–76.
- Kiran S, Thompson CK. The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language, and Hearing Research* 2003;46:773–787.
- Kiran S, Thompson CK, Hashimoto N. Training grapheme to phoneme conversion in patients with oral reading and naming deficits: A model based approach. *Aphasiology* 2001;15:855–876.
- Koul R, Corwin M, Hayes S. Production of graphic symbol sentences by individuals with aphasia: Efficacy of computer-based augmentative and alternative communication intervention. *Brain and Language* 2005;92:58–77. [PubMed: 15582036]
- Lustig AP, Tompkins CA. A written communication strategy for a speaker with aphasia and apraxia of speech: Treatment outcomes and social validity. *Aphasiology* 2002;16:507–521.
- Maas E, Barlow J, Robing D, Shapiro L. Treatment of sound errors in aphasia and apraxia of speech: Effects of phonological complexity. *Aphasiology* 2002;16:609–622.
- Martin N, Fink R, Laine M. Treatment of retrieval deficits with contextual priming. *Aphasiology* 2004;18:457–471.
- Mayer JF, Murray LL. Approaches to the treatment of alexia in chronic aphasia. *Aphasiology* 2002;16:727–743.
- Murray LL, Karcher L. A treatment for written verb retrieval and sentence construction skills. *Aphasiology* 2000;14:585–602.
- Peach RK, Wong PCM. Integrating the message level into treatment for agrammatism using story retelling. *Aphasiology* 2004;18:429–441.
- Raymer A, Ellsworth TA. Responding to contrasting verb retrieval treatments: A case study. *Aphasiology* 2002;16:1031–1045.
- Raymer AM, Rowland L, Haley M, Crosson B. Nonsymbolic movement training to improve sentence generation in transcortical motor aphasia: A case study. *Aphasiology* 2002;16:493–506.
- Renvall K, Laine M, Laakso M, Martin N. Anomia treatment with contextual priming: A case study. *Aphasiology* 2003;17:305–328.
- Renvall K, Laine M, Martin N. Contextual priming in semantic anomia: A case study. *Brain and Language*. in press
- Rose M, Douglas J, Matyas T. The comparative effectiveness of gesture and verbal treatments for a specific phonologic naming impairment. *Aphasiology* 2002;16:1001–1030.
- Schneider S, Thompson CK. Verb production in agrammatic aphasia: The influence of semantic class and argument structure properties on generalisation. *Aphasiology* 2003;17:213–241.
- Simmons-Mackie NN, Kearns KP, Potechin G. CAC Classics. Treatment of aphasia through family member training. *Aphasiology* 2005;19:583–593.
- Spencer KA, Doyle PJ, McNeil MR, Wambaugh JL, Park G, Carroll M. Examining the facilitative effects of rhyme in a patient with output lexicon damage. *Aphasiology* 2000;14:567–584.
- Thompson CK, Shapiro LP, Kiran S, Sobecks J. The role of syntactic complexity in treatment of sentence deficits in agrammatic aphasia: The complexity account of treatment efficacy. *Journal of Speech, Language, and Hearing Research* 2003;46:591–607.
- Wambaugh J. A comparison of the relative effects of phonologic and semantic cueing treatments. *Aphasiology* 2003;17:433–441.
- Wambaugh J, Cameron R, Kalinyak-Fliszar M, Nessler C, Wright S. Retrieval of action names in aphasia: Effects of two cueing treatments. *Aphasiology* 2004;18:97–1004.
- Wambaugh JL, Linebaugh CW, Doyle P, Martinez AL, Kalinyak-Fliszar M, Spencer KA. Effects of cueing treatments on lexical retrieval in aphasic speakers with different levels of deficit. *Aphasiology* 2001;15:933–950.
- Wambaugh JL, Martinez AL. Effects of modified response elaboration training with apraxic and aphasic speakers. *Aphasiology* 2000;14:603–617.

Youmans G, Holland A, Munoz ML, Bourgeois M. Script training and automaticity in two individuals with aphasia. *Aphasiology* 2005;19:435–450.

References

- Basso A, Capitani E, Vignolo L. Influence of rehabilitation of language skills in aphasia patients. A controlled study. *Archives of Neurology* 1979;36:190–196. [PubMed: 426663]
- Bernstein-Ellis, E.; Wertz, RT.; Shubitowski, Y. More pace, less fillers: A verbal strategy for a high-level aphasic patient. In: Brookshire, RH., editor. *Clinical aphasiology*. 17. BRK Publishers; Minneapolis, MN: 1987. p. 12-22.
- Bergin AE. Some implications of psychotherapy research for therapeutic practice. *Journal of Abnormal Psychology* 1966;71:235–246. [PubMed: 5941041]
- Boyle M, Coelho CA. Application of semantic feature analysis as a treatment for dysnomia. *American Journal of Speech-Language Pathology* 1995;4:94–98.
- Breuer, J.; Freud, S. *Studies on hysteria*. Basic Books; New York: 1957.
- Brookshire RH. Subject description and generality of results in experiments with aphasic adults. *Journal of Speech and Hearing Disorders* 1983;48:342–346. [PubMed: 6645428]
- Chambless DL, Hollon SD. Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology* 1988;66:7–18. [PubMed: 9489259]
- Connell PJ, Thompson CK. Flexibility of single subject experimental designs. Part III. Using Flexibility to design or modify experiments. *Journal of Speech and Hearing Disorders* 1986;51:204–214. [PubMed: 3525987]
- Gorman, BS.; Allison, CB. Statistical alternatives for single case designs. In: Franklin, RD.; Allison, DB.; Gorman, BS., editors. *Design and analysis of single case research*. Lawrence Erlbaum Associates; Mahway, NJ: 1996. p. 159-214.
- Holland AL, From DS, DeRuyter F, Stein M. Treatment efficacy: Aphasia. *Journal of Speech and Hearing Research* 1996;39:S27–S36. [PubMed: 8898264]
- Kazdin, AE. *Single case research designs*. Oxford University Press; New York: 1982. p. 230-261.
- Kearns KP. Flexibility of single subject designs. Part II. Design selection and arrangement of experimental phases. *Journal of Speech and Hearing Disorders* 1986;51:204–213. [PubMed: 3525987]
- Kearns, KP.; Thompson, CK. Technical drift and conceptual myopia: The Merlin Effect. In: Prescott, TE., editor. *Clinical aphasiology*. 19. BRK Publishers; Minneapolis, MN: 1991. p. 31-40.
- Kiran S, Thompson CK. The role of semantic complexity in treatment of naming deficits: Training semantic categories in fluent aphasia by controlling exemplar typicality. *Journal of Speech, Language, and Hearing Research* 2003;46:773–787.
- Le Dorze G, Jacob A, Coderre L. Aphasia rehabilitation with a case of agrammatism: A partial replication. *Aphasiology* 1991;5:63–85.
- Lincoln N, McGuirk E, Mullley G, Lendrem W, Jones A, Mitchell J. Effectiveness of speech therapy for aphasic stroke patients: A randomized controlled trial. *Lancet* 1984;1:1197–1200. [PubMed: 6202993]
- McReynolds, LV.; Kearns, KP. *Single subject experimental designs in communicative disorders*. University Park Press; Baltimore: 1983.
- McReynolds LV, Thompson CK. Flexibility of single subject experimental designs. Part I. Review of the Basics of Single Subject Design 1986;51:194–203.
- Renvall K, Laine M, Martin N. Contextual priming in semantic anomia: A case study. *Brain and Language*. in press
- Schuell, H.; Sefer, J. A year of aphasia therapy: A case study. In: Schuell, HS., editor. *Aphasia theory and therapy: Selected lectures and papers of Hildred Schuell*. University Park Press; Baltimore: 1974. p. 157-172.
- Thompson, CK. Aphasia treatment research: the state of the science; Paper presented at the International Aphasia Research Conference; 2000.

- Thompson, CK.; Kearns, KP. An experimental analysis of acquisition, generalization, and maintenance of naming behavior in a patient with anomia. In: Brookshire, RH., editor. Clinical aphasiology conference proceedings. BRK Publishers; Minneapolis, MN: 1981.
- Thompson, CK.; Kearns, KP. Analytical and technical directions in applied aphasia analysis: the Midas Touch. In: Prescott, TE., editor. Clinical aphasiology. 19. BRK Publishers; Minneapolis, MN: 1991. p. 41-54.
- Thompson CK, Kearns KP, Edmonds. An experimental analysis of acquisition, generalization, and maintenance of naming behavior in a patient with anomia. *Aphasiology*. in press
- Vignolo L. Evolution of aphasia and language rehabilitation: A retrospective exploratory study. *Cortex* 1964;1:344–367.
- Wambaugh J. A comparison of the relative effects of phonologic and semantic cueing treatments. *Aphasiology* 2003;17:433–441.
- Wertz RT, Weiss D, Aten J, Brookshire R, Garcia-Bunuel L, Holland AL, et al. Comparison of clinic, home, and deferred language treatment for aphasia: A Veterans Administration cooperative study. *Archives of Neurology* 1986;43:653–658. [PubMed: 3524513]

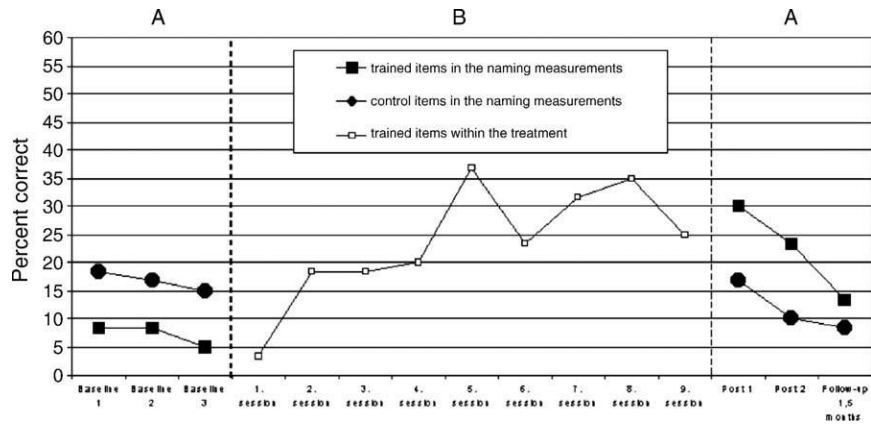


Fig. 1. ABA design strategy used by Renvall et al. (in press) to examine the effects of contextual priming on naming in a patient with anomia. Naming performance on trained and control items during A phases of the study and that on trained items during the B (treatment) phase is shown. Reprinted with permission.

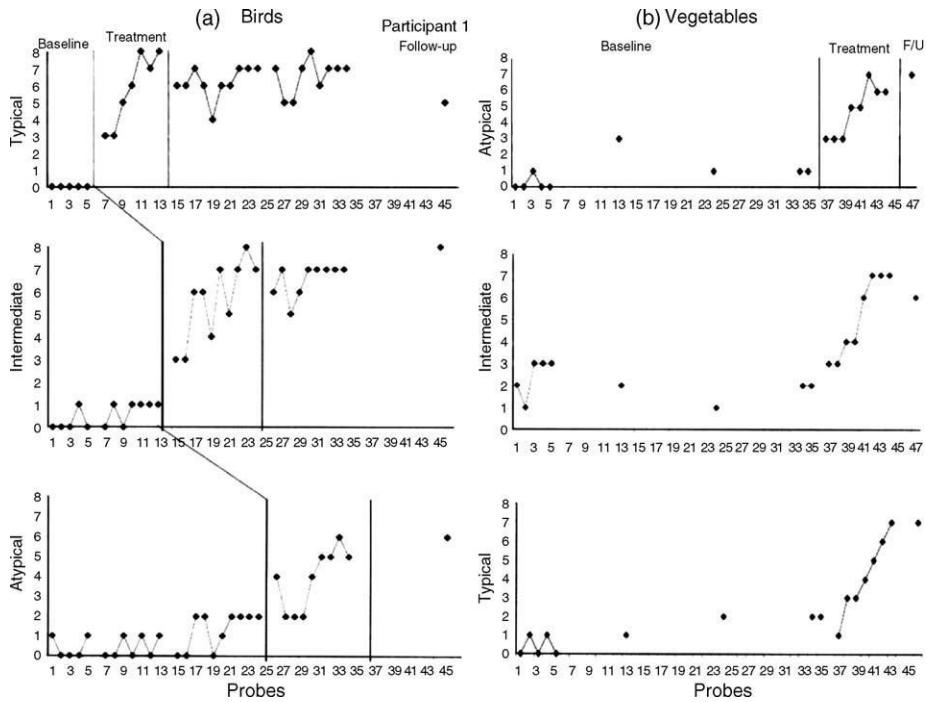


Fig. 2. Multiple baseline design used by Kiran and Thompson (2003) to investigate the effects of semantic complexity in treatment of naming in fluent aphasic participants. Naming performance on typical, intermediate, and atypical items for the category *birds* and that on atypical, intermediate, and typical items for the category *vegetables* is shown for Participant 1 in the study. Reprinted with permission.

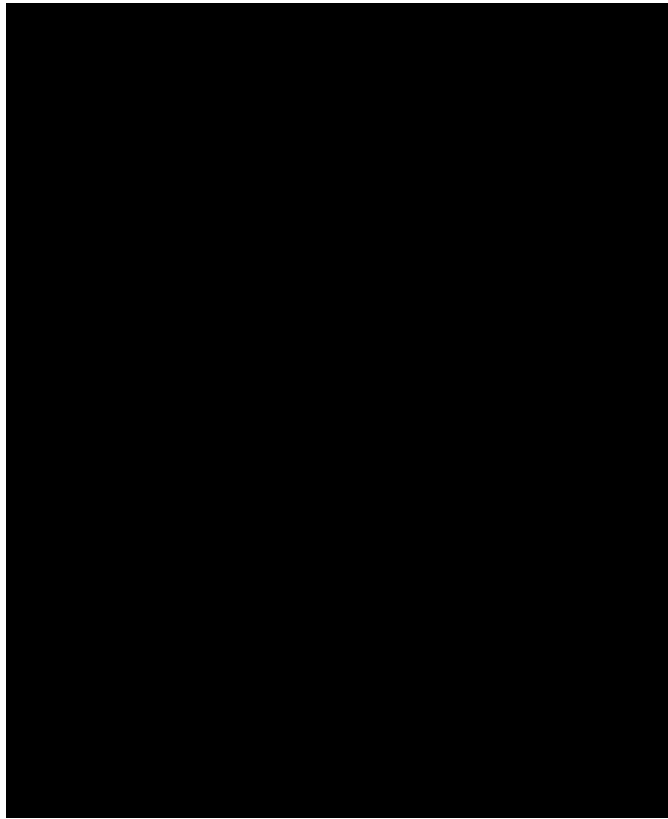


Fig. 3. Alternating treatment design, combined with a multiple baseline design across behaviors, used by Wambaugh (2003) to examine the effects of phonologic cueing (PCT) vs. semantic cueing treatment (SCT) in a single patient with chronic anomia. Percentage correct naming of word sets is shown across phases of the study. Reprinted with permission.

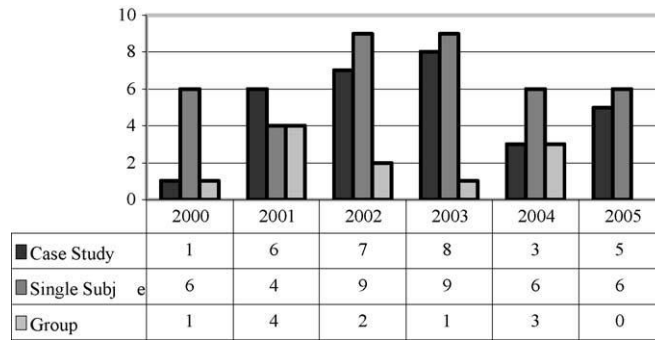


Fig. 4. Aphasia treatment studies by design type published from 2000 to 2005.

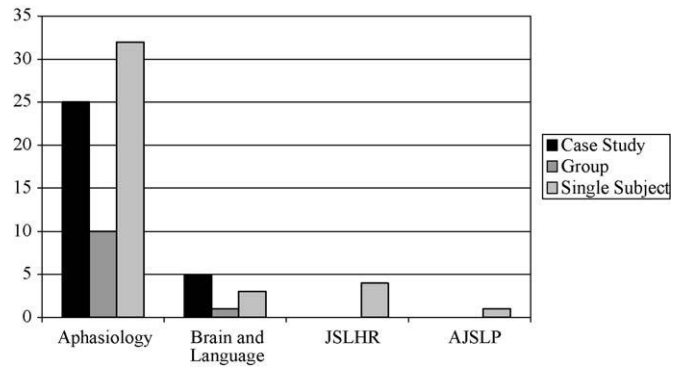


Fig. 5. Aphasia treatment studies published by journal from 2000 to 2005.

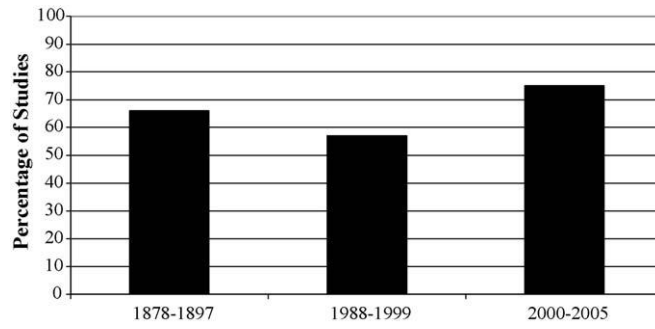


Fig. 6. Published single subject controlled experimental studies demonstrating internal validity in 1978–1987, 1988–1999, and 2000–2005 reviews of the literature.

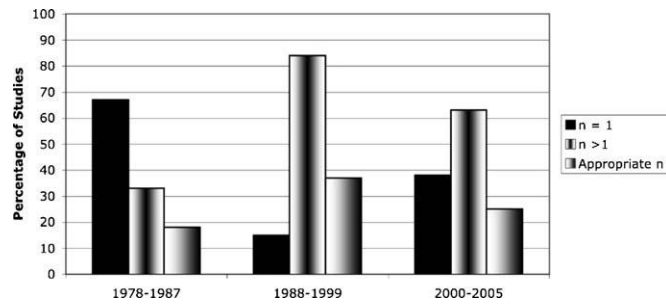


Fig. 7. Number of participants in published single subject controlled experiments in 1978–1987, 1988–1999, and 2000–2005 reviews of the literature.

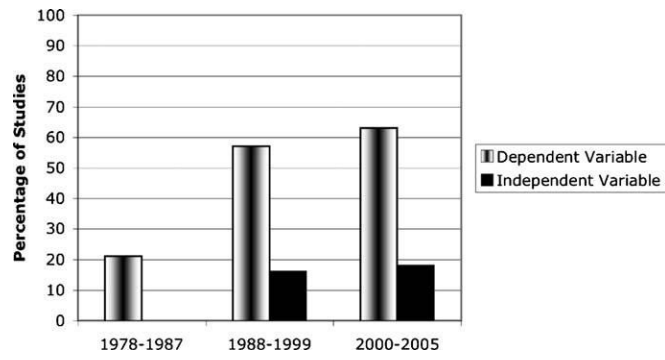


Fig. 8. Published single subject controlled experiments reporting reliability data in 1978–1987, 1988–1999, and 2000–2005 reviews of the literature.