Design Issues in Treatment Efficacy Research for Child Language Intervention: A Review of the Literature

Méthodes utilisées pour les études sur l'efficacité des interventions en orthophonie et en audiologie auprès des enfants : survol de la littérature

Patricia L. Cleave
Dalhousie University, Halifax, Nova Scotia

Abstract

Establishing the effects of speech and language services is an issue of great importance for speech-language pathologists. In the past decade, a variety of research designs have been used in treatment efficacy studies. Each of these have their strengths and weaknesses. Recognizing this is critical if clinicians are to appropriately apply information from treatment efficacy studies. In this paper, efficacy studies of intervention for child language impairments which have been published over the last decade are reviewed and evaluated in light of design issues.

Abrégé

L'établissement de l'incidence des services d'orthophonie et d'audiologie revêt une grande importance pour les orthophonistes et les audiologistes. Au cours de la dernière décennie, on a utilisé différentes méthodes de recherche pour effectuer des études sur l'efficacité des traitements. Chacune de ces méthodes comportait des forces et des faiblesses. Les cliniciens doivent en tenir compte s'ils veulent utiliser l'information de ces études de manière appropriée. Le présent article passe en revue les études sur l'efficacité des interventions auprès des enfants souffrant d'un trouble du langage qui ont été publiées au cours des dix dernières années. Il les examine du point de vue des méthodes utilisées.

Key words: treatment efficacy, child language intervention, research design

Establishing the effects of speech and language services is an issue of great importance for speech-language pathologists. For the practising clinician, the information from treatment efficacy studies is an important part of the clinical decision making process. However, applying the information from these studies is rarely straightforward. The purpose of this paper is to provide a survey of the literature in order to identify important factors when evaluating efficacy research. Further, this paper examines recent efficacy studies on intervention for child language impairments in light of these factors.

There are a variety of ways in which treatment efficacy studies are conducted, each with its own particular strengths and weaknesses. Given the complexity of the area of development being studied and the challenges involved in research with human participants it is unlikely, if not impossible, that the perfect efficacy study of child language intervention could be conducted. Thus, being aware of the strengths and weaknesses of a particular study is important if one is to appropriately evaluate and apply the results of efficacy research. In this article, efficacy studies will be reviewed primarily from a design perspective. Three broad areas will be considered: experimental control, clinical significance, and clinical feasibility. For recent reviews of treatment efficacy research of language intervention with children which focus more on the outcomes of the studies, the reader is referred to Law (1997) and Law, Boyle, Harris, Harkness, and Nye (1998).

A number of papers discussing the issues involved in treatment efficacy research and stressing the need for such research were written approximately 10 years ago (e.g., Butler, 1990; Ellis Weismer, 1991; Fey, 1990; Goldstein, 1990; Kluppel, 1991). Furthermore, the American Speech-Language-Hearing Foundation sponsored a Conference on Treatment Efficacy in 1989 whose proceedings were published as an edited volume (Olswang, Thompson, Warren, & Minghetti, 1990) and a clinical forum was published in the journal, *Language, Speech, and Hearing Services in Schools* in October 1991. More recently, the focus has been on treatment outcomes re-

search as evidenced by things such as the Clinical Forum on treatment outcomes in *Language, Speech, and Hearing Services in Schools* in October 1998 and the development of the National Outcomes Measurement System (NOMS) by the American Speech-Language-Hearing Association (ASHA). In addition, the Canadian Association of Speech-Language Pathologists and Audiologists (CASLPA) held a meeting on outcome measures before their annual convention in May 2000.

Olswang (1998) places efficacy research and outcomes research at the two ends of a continuum of treatment research. Treatment efficacy research is concerned with proving the benefits of treatment and, therefore, demands well controlled experimental manipulations and data collection. These experimental controls mean that treatment efficacy research not only provides information on the impact of an intervention, but it also provides information that is important for furthering our understanding of language development and language disorders. Outcomes research seeks to identify the benefits of treatment in the "real world" conditions of clients' lives; that is, as it interacts with everything else occurring in a person's life. In the present paper, the focus is on articles at the treatment efficacy end of the continuum—those studies which seek to prove the effects of treatment. In the present article, a review of a number of journals that typically contain articles dealing with language impairments in children was undertaken. The journals of key professional associations were reviewed. These included the *Journal of Speech-Language Pathology and Audiology*, the journal of CASLPA; the *Journal of Speech, Language, and Hearing Research*, the *American Journal of Speech-Language Pathology*, and *Language, Speech and Hearing Services in Schools*, all journals of ASHA; and the *International Journal of Language and Communication Disorders* (previously titled the *British Journal of Disorders of Communication* and *European Journal of Disorders of Communication*), the journal of the Royal College of Speech and Language Therapists. In addition, *Child Language Teaching and Therapy* was reviewed because of its focus on language intervention with children. These journals were surveyed for the years 1990 - 1999 for all articles that reported on investigations of interventions with children with language impairments, with or without cognitive delays. Reports dealing with children with diagnoses which fell on the Pervasive Developmental Disorder (PDD) spectrum or those using AAC systems were not included. Further, interventions directed toward written language development were not included because most of the articles reporting on these would be found in other journals (e.g., *The Reading Teacher, Reading Research Quarterly, Journal of Experimental Child Psychology*). Also excluded were descriptions of interventions that included no or very limited data on the effects of the intervention. These were articles whose main purpose was to describe an intervention program. However, case studies, whose main purpose was to describe how an individual responded to treatment, were included. A total of 52 articles were identified during the review.

Experimental Design

The vast majority of the studies reviewed report that treatment was successful in improving the participating children's language abilities. The strength with which the claim can be made, however, is dependent largely upon the design of the study. In particular, it is important to examine how experimental control was achieved in order to rule out other possible explanations for the changes noted. Studies were classified as involving a group design, single subject experimental design, or case study. The breakdown by journal and type of study is shown in Table 1.

Group Design Studies

Studies in which the results were aggregated across subjects were classified as group designs. Experimental control could be achieved using either a between group or within group comparison. Twenty-five studies fell into the group design category. A variety of between group comparisons were involved. Some studies used a standard no-treatment control group (e.g., Gibbard, 1994; McDade & McCarten, 1998; Ward, 1999) in which no services were provided. This design does provide experimental control but concerns regarding the ethics of not providing treatment have been raised. In other studies, the control group received the typical clinical or educational services (e.g., Best, Melvin, & Williams, 1993; Kaufmann, Prelock, Weiler, Creaghead, & Donnelly, 1994; Masterson & Perrey, 1999). Although this design avoids some of the ethical issues that arise with no treatment control groups, it does remove a degree of experimental control. There is no control over the treatment the control group received and generally few details are given or steps taken to ensure uniformity of treatment. This problem is compounded because there is always the possibility that the standard clinical services also address the same area of language as the experimental treatment, which would mitigate against finding a significant effect of the experimental treatment. Obviously, if the experimental treatment is found to result in greater improvement than typical services, this is not a concern.

As an alternative to the no-treatment control group, a few studies utilized a delayed-treatment control group where the control group received the experimental intervention after

serving as controls (e.g., Fey, Cleave, Long, & Hughes, 1993; Girolametto, Pearce, & Weitzman, 1996; Robertson & Ellis Weismer, 1999). Similar ethical issues arise when temporarily withholding treatment from clients. However, given that all participants eventually receive the experimental treatment, the concerns related to this design are mitigated somewhat. There is a particular advantage to this design when the treatment results from the delayed treatment are reported (e.g., Fey et al., 1993). When it is demonstrated that the control group responds to the intervention under investigation in a similar manner as the experimental group, this is evidence that the differences between the two groups were not the result of the control group being more resistant to change. In addition, the results of intervention with the control group can serve as a replication of the initial results.

With both the no-treatment and delayed-treatment control designs there is the possibility that general contact with the experimenter rather than the specific intervention program may have been contributing to or even be responsible for the changes seen in the experimental group. An attempt to control for this could be done by having the control group participants attend a similar number of sessions but without the experimenter using any particular intervention techniques. This would be comparable to the 'sugar pill' in drug trials. None of the studies reviewed included this control.

As noted, ethical issues are raised in all three of the above designs given that, in each one, a particular treatment is denied or delayed. Recently, the three major Canadian research granting agencies (i.e., Medical Research Council, Natural Sciences and Engineering Research Council and the Social Sciences and Humanities Research Council) published the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans* (1998). This policy statement lays out requirements for ethical conduct and review for all individuals and institutions (e.g., universities, hospitals) which receive funding from these bodies. Thus, virtually all research on communication disorders conducted in Canada is governed by these guidelines. The policy statement covers any research involving humans. It addresses issues central to all types of research such as the requirement for free and informed consent and confidentiality. Of particular interest to this discussion is the section devoted to clinical trials. Although the clinical trials guidelines are written with an emphasis on pharmaceutical trials, most of the issues also apply to treatment efficacy research in communication disorders. The Tri-Council policy statement includes a discussion of placebo-controlled studies which has relevance for the designs described above. Under these guidelines, placebo-controlled studies are ethical under certain conditions. Examples of these conditions include when there is no standard treatment, when effective treatment is not available to patients

Table 1. Treatment Efficacy Articles by Journal and Design Type.

Journal	Group Design	Single Subject Design	Case Study	Total Number of Articles
Journal of Speech-Language-Hearing Research	9	8		17
Child Language Teaching and Therapy	4	3	9	16
Language, Speech, and Hearing Services in Schools	3	4		7
British/European/International Journal of Disorders of Communication	6	1		7
American Journal of Speech-Language Pathology	3	1		4
Journal of Speech-Language Pathology and Audiology		1		1

due to cost constraints or short supply, and when testing an add-on to the standard treatment.

The studies which involved providing typical services for the control group (e.g., Hess, Wagner, DeWald, & Cohn, 1993; Masterson & Perrey, 1999) clearly meet ethical guidelines. Of the studies employing no treatment and delayed treatment designs, two involved young children for whom a definitive diagnosis of impairment could not be made and for which no standard treatment exists (i.e., Robertson & Ellis Weismer, 1999; Ward, 1999). Others took advantage of the fact that there were waiting lists for services. Gibbard (1994) explicitly states this but it was also the case for the study reported by Fey and his colleagues (Fey, Cleave, Long, & Hughes, 1993; Fey et al., 1997). McDade & McCarten (1998) used children who could not attend the treatment sessions as their control group. Although this raises concerns about nonrandom assignment, to be discussed later, it does address the ethical issues of withholding treatment.

Instead of making comparisons to a no- or delayed-treatment control group, a number of studies compared different intervention programs. These studies differed from studies in which the control group received the standard clinical or educational programs in that the experimenter designed and implemented both treatment programs. Thus, the nature of the intervention with both groups was known. In addition, this design avoids the problem that arises when there is no contact with the control group, as discussed earlier. This design is most appropriate when a new intervention is being compared to one that has been proven effective. If this is not the case, difficulties arise when both programs appear to be beneficial. There is no evidence that the gains seen are greater than those which would have taken place with no treatment. Thus, neither treatment has been shown to be efficacious. This is less of a problem if a significant difference is found between the two programs. In this case, the assumption that the better of the two programs is more effective than no treatment is appropriate. This difficulty can be avoided by including a delayed-treatment control group in addition to the two treatment groups as seen in the study by Fey and his colleagues (Fey et al., 1993, 1997).

The intervention programs compared in the studies reviewed for this article differed along a number of dimensions. A number of studies contrasted two different approaches to therapy such as Milieu Therapy versus the Communication Training Program for learning early word combinations (Yoder, Kaiser, & Alpert, 1991) or modelling versus self-evaluation for learning the use of constraint questions (Mann-Mandelbaum, 1990). Some studies involved programs differing in intervention agents, that is who provides the treatment (e.g., Fey et al., 1993; Gibbard, 1994; Kot & Law, 1995). Others compared individual to classroom-based treatment (e.g., Roberts, Prizant, & McWilliam, 1995; Wilcox, Kouri, & Caswell, 1991). Finally, there were studies in which the control group received therapy targeting another language area (e.g., Hyde Wright, 1993) or another developmental domain such as cognition (e.g., Gibbard, 1994).

There were three studies in which group results of within subject comparisons were presented. In within subject treatment efficacy designs, each participant receives both interventions, each targeting different language goals. However, rather than examining progress on an individual participant level, as is done in single subject designs, the data for each experimental condition are combined from all participants and group comparisons of the effects of the two treatments are made. When within subject designs are used, since each person receives both interventions, factors such as parental interaction style and personal history are equated across the two treatment conditions. However, there is always the possibility that the findings are at least partly the result of an interaction between the two interventions. In this review, two studies (Camarata, Nelson, & Camarata, 1994; Nelson & Camarata, 1996) assigned specific language targets for each child to either recasting or imitation treatment conditions. Within subject comparisons were made between targets treated through recasting and those assigned to the imitation condition. In a unique design, Warren and his colleagues published a study in which they took data collected through a number of single subject design studies of milieu therapy and aggregated them to do group comparisons between baseline and treatment scores (Warren, Gazdag, Bambara, & Jones, 1994).

Group designs have the advantages of being able to employ statistical procedures to assess the reliability of the findings and of allowing generalization to the population from which the participants were chosen. Generalization, however, requires a number of assumptions about the nature of the groups, such as the assumptions that the variability in the two groups is equal or that all participants will respond in a similar manner to the intervention(s) (Silverman, 1998). In 1994, Attanasio published an article in which he outlines some of the difficulties with group designs for efficacy studies. Random assignment to groups, sometimes with matching of subjects on critical variables, is used to try to ensure the groups are equal across a wide range of variables both measured and unmeasured. Attanasio (1994) pointed out that random as-

signment to groups is critical to the assumptions on which generalization is based and stated that true randomization is often not done in efficacy studies in our field. Of the studies reviewed here, 17/25 used random assignment to groups. Other studies used only matched subjects (e.g., Wing, 1990; Hyde Wright, 1993), while others used children from naturally formed groups (e.g., from another class in the same school – Kaufmann et al., 1994; Hess et al., 1993), from children who could not attend the experimental sessions (McDade & McCarten, 1998), or whose parents' failed to respond to the offer of additional services (Masterson & Perrey, 1999). Although researchers can test to see that the groups did not differ on pretest measures, this does not ensure that they did not differ on untested, but important variables. Therefore, when studies do not use random group assignment, caution must be taken when generalizing the results. The necessary degree of caution depends on the nature of the groups. If the naturally occurring groups were randomly (or semi-randomly) created (e.g., another class), the concern is less than if the groups were divided based on a potentially critical factor (e.g., parents not responding to the offer of extra services which may suggest less support at home).

Another weakness of group designs that Attanasio (1994) discusses is that the aggregated data only provide information on the central tendency of the group, not on individuals. It is certainly possible, if not likely, that although the treatment was effective on a group basis, not all individuals responded to the treatment. In group design studies, if the researchers provide data on individual participants in addition to the group analyses, the reader is able to judge how representative the group results are for the individuals. Twelve of the group design studies reviewed provided some individual data on treatment effects.

A final factor that is important to consider with group designs is the number of participants. The adequacy of the sample size is determined by a number of factors including the purpose of the study, the expected size of the treatment effect, and the variability of the data (Schiavetti & Metz, 1997). The number of participants involved is an important factor in determining the statistical power of a study. All things being equal, the smaller the group size, the less the power. When statistical procedures are used without sufficient power to find a group difference, the possibility of a Type II error (deciding there is no group difference when one truly exists) is high. Sample sizes in the studies reviewed ranged from 8 (4 per group) (Hyde Wright, 1993) to 58 (28 and 30 per group) (Yoder & Warren, 1998). Across the 25 studies, the number of participants was fairly evenly distributed between the two extremes. Silverman (1998) states that more than 10 participants are usually required for group designs. Only one of the studies reviewed did not meet this minimum criteria (Hyde Wright, 1993).

Single Subject Design Studies

Eighteen of the studies identified involved single subject designs. These involved a variety of procedures designed to establish experimental control. Three studies utilized a simple baseline period as their experimental control. There were two (Tyler & Sandoval, 1994), three (Schuele, Rice, & Wilcox, 1995) and five (Biggs Masters & Pine, 1992) measurements taken during baseline. By obtaining multiple measures before the treatment program is initiated, the stability of the behaviour under examination can be established. Thus, if a difference is seen in measurements taken after treatment, it is assumed not to be just random fluctuation but a true effect of treatment. The number of measurements needed to establish a stable baseline depends on the variability of the behaviour under examination, but generally more than two or three are required unless the behaviour is unusually stable. What a simple baseline design does not control for is the possibility that something else responsible for the change happened coincidently with the initiation of treatment.

Other single subject experimental design studies have achieved greater control by using multiple baselines lasting different lengths of time for each participant (e.g., Ezell & Jarzynka, 1996; Kaiser & Hester, 1994; Venn, Wolery, Fleming, DeCesare, Morris, & Cuffs, 1993). If the effect of treatment is seen in each participant even though treatment was initiated at different times, the possibility that another factor that occurred coincidently at the beginning of treatment was responsible is greatly reduced. In the studies employing this design, baselines ranged from a minimum of three to a maximum of five in one study (Yoder, Spruyterburg, Edwards, & Davies, 1995) to a minimum of five and a maximum of 40 in another (Warren, Yoder, Gazdag, Kim, & Jones, 1993). An adequate baseline or baselines is critical for experimental control in single subject designs. However, when baselines last for a substantial length of time, the ethical issues of delaying treatment, which were discussed earlier, are raised.

A number of studies involved an alternating treatment design where each participant received both treatment protocols. To judge the relative effect of the treatments, specific vocabulary (Haynes, 1992; Sorensen & Fey, 1992), semantic relations (Clements-Baartman & Girolametto, 1995; Kim & Lombardino, 1991) or storybooks (Bradshaw, Hoffman,

& Norris, 1998) were assigned to the different treatment conditions. With this design, the assumption is that other factors, such as maturation, would affect treatment targets in both intervention protocols the same and, thus, if one treatment leads to greater gains, those gains must reflect effects of that treatment.

As with group designs, there are strengths and weaknesses to single subject designs and it is important to interpret the findings in light of these. A major advantage with single subject designs is that the data from individuals are available rather than just aggregate data. The reader can see how a real person responded to the treatment. Further, one can determine how many participants did not respond to the intervention. The weakness, of course, is that generalizability of the results is limited. The more participants involved in the study and the more consistent the findings across participants, the more confidence one can have that others will respond in a similar manner. The number of participants in the studies reviewed ranged from two (Ezell & Jarzynika, 1996; McGregor, 1994) to eight (Biggs Masters & Pine, 1992; Goldstein, English, Shafer, & Kaczmarek, 1997).

In addition to replications within a study, there can be replications of an intervention program or technique across studies. In this review, a total of five were single subject design studies of milieu therapy (Kaiser & Hester, 1994; Venn et al, 1993; Warren et al., 1993; Yoder et al., 1994; Yoder et al., 1995). Further, as noted previously, Warren and his colleagues (1994) published a study in which the results from a number of single subject studies were combined and group analyses done. An alternative which was not found in this review is a meta-analysis involving a number of related single subject research studies (Busk & Serlin, 1992).

Beyond the difficulties with generalization, some traditional assumptions of single subject designs can limit their appropriateness for certain language intervention research. Under these designs, the experimental treatment is presumed to be responsible for the changes seen when those changes are tied in time to the introduction of the intervention. If the impact of intervention is delayed, the claim that the intervention caused the changes is less strong (Kratochwill, 1992). However, depending on one's theory of language acquisition or how the intervention is designed to impact language, immediate effects may not be expected. Under these conditions, single subject designs may be inappropriate. Group designs may be a better choice since demonstrating the effects of treatment does not rely on tying the observed changes to the introduction of intervention. Further, Levin (1992) cautions that when the treatment effects are very broad or impact individuals who differ on a variety of dimensions, this also raises doubts about the inference that the intervention was responsible for the changes noted. On the other hand, it is often clinically desirable to impact development broadly. Still, interventions which impact behaviours or situations to which there is no conceptual or theoretic link lack discriminant validity which weakens the claim that the intervention was responsible (Levin, 1992). Thus, it is important that the researchers make clear predictions of what the impact(s) of the intervention will be based on their theories of language and language learning.

Discriminant validity can be demonstrated through the use of multiple baselines from each participant (Levin, 1992). In this case, the researcher is monitoring a number of behaviours or goals. Treatment goals are treated directly and expected to change with the introduction of the treatment. Control goals, on the other hand, are not treated and are not expected to be impacted by the treatment. However, they are assumed to be influenced by all other factors (e.g, maturation, attention from the examiner, etc.) in the same manner as the experimental goals. Thus, if growth is seen on the experimental goals but not on the control goals, it is assumed that the treatment was responsible and discriminant validity is shown. A study may also involve other behaviours which are expected to be indirectly affected by treatment, possibly with a time lag. These are referred to as generalization goals. The identification of appropriate control and generalization goals is very much dependent on one's theory of language and language learning. For example, as Long and his colleagues (Long, Olswang, Brian, & Dale, 1997) noted, there are two views of how children move from single words to two word combinations. One is that there is a lifting of a general constraint such that two word utterances of a variety of sorts are possible. The other view proposes that children's knowledge of abstract grammatical classes and semantic roles allows them to build specific semantic relations. The first theory would predict that the child is learning to combine two words in general. Thus, regardless of what two word utterances were used in treatment, any other two word utterance would be seen as a generalization goal since treatment would be expected to impact that learning. The second would predict that learning would occur on a semantic relation by semantic relation basis. Thus, if action-object and agent-action utterances were trained, other action relations (e.g., action-location) would be viewed as generalization goals while two word utterances that mark semantic relations such as possessor-possession would be viewed as control goals.

Of the studies reviewed, only three had direct measures for discriminant validity. Two studies monitored the children's acquisition of untreated items as control goals. The control goals in the studies reviewed were untreated exemplars of the target set such as specific words in a word learning study (e.g., Sorensen & Fey, 1992) and semantic relations in a study of an intervention targeting the development of semantic relations (e.g., Clements-Baartman & Girolametto, 1995). Conversely, studies which were designed to teach strategies for word finding (McGregor, 1994) or comprehension of jokes (Ezell & Jarzynka, 1996) used untrained items as generalization goals. The third study which documented discriminant validity involved the collecting baselines of behaviours not expected to be affected by a given intervention. In the study by Ezell & Goldstein (1991), children were trained to use requests for clarification in response to four types of communication breakdowns. Training was introduced for the second type of inadequate message (e.g., interfering signal) when mastery was reached on the first (e.g., unfamiliar word). Training on the third type followed mastery on the second and so on. The children's responses to all four types of breakdowns were measured throughout the study. Thus, the researchers were able to show that each training procedure impacted only the type of inadequate message it targeted.

As stated earlier, some studies have involved an alternating treatment design in which each participant received both interventions. This design assumes that the children receiving the interventions do not actively apply information or skills learned in one intervention to targets of the second intervention. In this review, the studies that used an alternating treatment design compared the presentation of discrete items such as words (Sorensen & Fey, 1992), or semantic relations (Clements-Baartman & Girolametto, 1995; Kim & Lombardino, 1991), or storybooks (Bradshaw et al., 1998) under the two treatment regimes lasting a relatively short time. As such, the possibility of interaction between the two treatment protocols may have been slight. However, this type of design clearly is not appropriate if the interventions are designed to teach the children rules or strategies which they can then apply, or to activate the children's language learning mechanism so that they can benefit more from the language surrounding them throughout the day.

Case Studies

Nine of the reviewed articles were classified as case studies. A study was placed in this category if only pre- and post-treatment data were provided. In other words, there was no experimental control to eliminate alternative influences, such as maturation, as factors responsible for the results. In three of the studies, there were treated and untreated targets. One study involved the acquisition of the regular past tense morpheme (Eyer & Leonard, 1994) and the other two studies addressed word finding skills (Easton, Seach, & Easton, 1997; Wittmann, 1996). For this report, the untreated words in these studies were viewed as generalization rather than true control goals as the treatment involved learning a rule or strategies which the child could apply to untreated words. Thus, these words did not provide experimental control. The category of case studies included reports on a single child (e.g. Klecan-Aker, 1993; Layton & Savino, 1990), as well as two reports of programs conducted with a group of children (Lamb, Bibby, & Wood, 1997; Sim, 1998).

Clinical Significance

When evaluating an efficacy study, in addition to evaluating the degree of experimental control, (i.e., how strong the evidence is that the intervention was responsible for the change noted), it is necessary to consider the clinical significance of that change. In other words, given that there was an effect of treatment, how important is it? There are a variety of parameters that should be examined when considering the clinical significance of a treatment efficacy study. As Attanasio (1994) points out, the statistical significance reported in a group design study does not address the importance or magnitude of the group difference. He suggests that group design studies should report the effect size in addition to the statistical significance level. Of the 25 studies in this review which incorporated group statistics, only four reported effect sizes. When single subject designs are used, because the scores of individual participants are reported in the article, the magnitude of the change achieved by each participant can be determined by examining the data directly. As an alternative to 'eyeballing' the data, Bain and Dollaghan (1991) discuss two metrics, the *Intervention Efficacy Index* (IEI) and the *Proportion Change Index* (PCI) as ways to quantify changes within an individual subject. Because both of these require age equivalent data, not all studies could employ them. However, only one study was found to have used the PCI. This was a group design study conducted by Yoder et al. (1991). No study reviewed used the IEI.

When evaluating the clinical significance of a treatment effect, it is important to have information on the maintenance of the new skills or, ideally, the continued improvement after treatment is discontinued. Of the 18 single subject design studies, seven included data from maintenance measures. These varied from measurements taken immediately after treatment

but without the supports provided by treatment (e.g., Venn et al., 1993) to 10 weeks post-treatment (e.g., Ezelle & Goldstein, 1991). The inclusion of maintenance data was less frequent in group designs but it did occur. For example, Hyde Wright (1993) reported on the children's performance one month post-treatment and the study by Ward (1999) involved examining children's outcomes three years after the provision of treatment. The study by Gottschalk et al. (1997) reported on a follow-up with the children from an earlier study of a treatment for metapragmatic development (Kaufmann, Prelock, Weiler, Creaghead, & Donnelly, 1994). Also, as part of their second report on their intervention project, Fey and his colleagues report on the development of 10 children five months after the completion of treatment (Fey, Cleave, & Long, 1997). Five out of the nine case studies included maintenance data. In these studies, the time after treatment ranged from one to two months (Sim, 1998) to one year (Layton & Savino, 1990).

Another important consideration is how broad the effect of treatment was. This can be viewed from a number of perspectives. Many of the single subject design studies included measures of generalization. These included generalization to another person (e.g., Yoder, Warren, Kim, & Gazdag, 1994); materials (e.g., Clements-Baartman & Girolametto, 1995); untrained items (e.g., McGregor, 1994) or setting (e.g., Warren, Yoder, Gazdga, Kim, & Jones, 1993). One study of peer interaction included not only measures of number of interactions but also examined sociometric ratings (Goldstein, English, Shafer, & Kaczmarek, 1997). Thus, the impact of treatment was documented in a related area. In group design studies, similar demonstrations of generalization can be seen. For example, Fey and his colleagues used language samples collected with the children's parents rather than the SLP as the data for their experimental measures (Fey et al., 1993; Fey et al., 1997) and Wilcox and her colleagues collected samples in the children's homes to evaluate a treatment that was provided at school (Wilcox et al., 1991). Robertson and Ellis Weismer (1999) collected parent report data on linguistic and social/behavioural development in addition to their direct linguistic measures. Furthermore, Ward (1999) looked at the number of children who were receiving services three years later as a measure of the impact of her treatment program. In contrast, some studies measured the impact of treatment in very narrow contexts (Haynes, 1993; Kim & Lombardino, 1991; Masterson & Perrey, 1999).

Of course, when evaluating the magnitude, permanence, and breadth of reported treatment effects, it is crucial that the nature of the intervention be considered. Some of the studies reviewed involved intervention programs designed to impact a relatively narrow area of language development and the effect of treatment was demonstrated in that narrow area. Studies of this type addressed areas such as the use of constraint questions (Mann-Mandelbuam, 1990), the use of requests for clarification (Ezelle & Goldstein, 1991), peer initiations (Schuele, Rice, & Wilcox, 1998) or the comprehension of jokes (Ezell & Jarzynka, 1996). These would not be expected to impact overall language development. However, others reported on intervention programs designed to impact language development more broadly and the measures used were consequently more comprehensive (e.g., Bradshaw et al., 1998; Fey et al., 1994; Kaiser & Hester, 1994; Ward, 1999).

Related to the nature of the treatment, is the length of intervention. Narrowly focussed treatments generally involve less treatment time than programs expecting to impact language broadly. When evaluating the clinical significance of a treatment effect, it is important to consider the cost of the treatment, in both terms of time and money, relative to the benefits. Of the studies reviewed, treatments varied from three sessions (Kaufmann et al, 1994) to six months (Yoder & Warren, 1998) or eight months (Fey et al., 1997). Obviously, a program that involved six or eight months of treatment would need to demonstrate much larger and broader effects than would a program lasting three sessions for those effects to be considered clinically significant.

Clinical Feasibility

The third group of features to consider when reading efficacy research addresses the issue of whether and how clinicians can incorporate the findings into their practice. If clinicians are to make use of the information from efficacy studies, sufficient detail in terms of the characteristics of the participants must be provided. The issue raised is whether there is enough information provided so that clinicians can determine if their clients are sufficiently similar to the experimental participants to expect similar results. In addition, it is critical that the intervention program(s) or techniques are described in enough detail to allow clinicians to implement them. Reports on interventions that were narrowly focussed generally provided sufficient intervention details (e.g., Hyde Wright, 1993; McGregor, 1994), but more broadly based interventions often contained few details regarding the intervention program(s) (e.g., Gibbard, 1994). The major reason for this is that there are often length limits for articles. It is difficult to provide sufficient details of both the intervention and results in a single article. As a means around this difficulty, Cleave and Fey (1997) published a separate article describing the in-

tervention programs involved in their study (Fey et al, 1993; 1994; 1997). Milieu therapy, the focus of a number of studies, has also been described in detail in a variety of places (e.g., Hart, 1985; Yoder & Warren, 1993). For some studies, providing the details of the intervention program(s) is less of an issue because they have implemented commercially available intervention programs, sometimes with a modification. Thus, the details of the intervention are available elsewhere (e.g., Girolametto et al., 1996).

Finally, if clinicians are to the adopt the new interventions, it is important that they be 'do-able' in a clinical setting. That is, they must involve reasonable demands, in terms of cost and expectations for the clinician, child, teacher, and/or parent. If significant changes to a program are needed before it can be implemented in a clinical setting, one cannot assume that the effects seen in the efficacy study will be replicated.

Conclusions

A variety of study designs has been used by researchers to examine the efficacy of language intervention with children. Each of these designs has its own strengths and weaknesses. They make different assumptions and have different requirements. Because of that, studies employing these designs serve different purposes and can make different claims. However, if conducted appropriately, they all can provide useful information in our development of effective, efficient intervention programs. Studies that provide adequate experimental control, either through group or single subject designs, allow us to have confidence in judging the effects of treatment. Case studies, while they do not involve the experimental control necessary to prove treatment effects, can be useful in identifying promising new procedures to test and in demonstrating extensions of tested protocols to new populations and settings. Studies of comprehensive treatment packages assess programs designed to address most, if not all, of children's typical language learning goals. However, studies of short-term therapies that affect a limited area of language can provide important information in the development of more comprehensive treatment programs. If we are to improve the services we provide to children with language learning needs, we need converging evidence of a treatment's effects from various types of studies and also replication of successful experiments with that treatment. In addition, it is vital that clinicians are able to evaluate the strengths and weaknesses of efficacy studies and make appropriate applications of the studies' findings to their clinical work.

Author Notes

The author wishes to thank Elizabeth Kay-Raining Bird for her helpful comments on an earlier draft of this article. Correspondence concerning this article should be addressed to Patricia L. Cleave, PhD, School of Human Communication Disorders, Dalhousie University, 5599 Fenwick Street, Halifax, Nova Scotia, B3H 1R2.

References

Attanasio, J. S. (1994). Inferential statistics and treatment efficacy studies in communication disorders. *Journal of Speech and Hearing Research, 37*, 755-759.

Bain, B. A., & Dollaghan, C. A. (1991). Treatment efficacy forum: The notion of clinically significant change. *Language, Speech and Hearing Services in Schools, 22*, 264-270.

Best, W., Melvin, D., & Williams, S. (1993). The effectiveness of communication groups in day nurseries. *European Journal of Disorders of Communication, 28*, 187-212.

Biggs Masters, J., & Pine, S. J. (1992). Incidental group language therapy: Verbal and preverbal children. *Child Language Teaching and Therapy, 8*,18-29.

Bradshaw, M. L., Hoffman, P. R., & Norris, J. A. (1998). Efficacy of expansions and cloze procedures in the development of interpretations of preschool children exhibiting delayed language development. *Language, Speech, and Hearing Services in Schools, 29*, 85-95.

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 187-212). Hillsdale, NJ: Erlbaum.

Butler, K. G. (1990). The future of language science: Impact on clinical practice. In C. Shewan (Ed.), *ASHA Reports 20: The "future of science and services" seminar* (pp. 51-58). Rockville, MD: ASHA.

Camarata, S. M., Nelson, K. E., & Camarata, M. N. (1994). Comparison of conversational-reeasting and imitative procedures for training grammatical structures in children with specific language impairment. *Journal of Speech and Hearing Research, 37*, 1414-1423.

Cleave, P. L., & Fey, M. E. (1997). Two approaches to the facilitation of grammar in children with language impairments: Rationale and description. *American Journal of Speech-Language Pathology, 6*, 22-32.

Clements-Baartman, J., & Girolametto, L. (1995). Facilitating the acquisition of two-word semantic relations by pre-schoolers with Down syndrome: Efficacy of interactive versus didactic therapy. *Journal of Speech-Language Pathology and Audiology, 19*,103-112.

Easton, C., Sheach, S., & Easton, S. (1997). Teaching vocabulary to children with wordfinding difficulties using a combined semantic and phonological approach: An efficacy study. *Child Language Teaching and Therapy, 13*, 125-142.

Ellis Weismer, S. (1991). Child language intervention: Research issues on the horizon. In J. Miller (Ed.), *Research on child language disorders: A decade of progress*, (pp. 233-243). Austin, TX: ProEd.

Eyer, J. E., & Leonard, L. B. (1994). Learning past tense morphology with specific language impairment: A case study. *Child Language Teaching and Therapy, 10*, 127-138.

Ezell, H. K., & Goldstein, H. (1991). Observational learning of comprehension monitoring skills in children who exhibit mental retardation. *Journal of Speech and Hearing Research, 34*, 141-154.

Ezell, H. K., & Jarzynka, M. A. (1996). An intervention for enhancing children's understanding of jokes and riddles. *Child Language Teaching and Therapy, 12,* 148-163.

Fey, M. E. (1990). Understanding and narrowing the gap between treatment research and clinical practice with language impaired children. In C. Shewan (Ed.), *ASHA Reports 20: The "future of science and services" seminar* (pp. 31-40). Rockville, MD: ASHA.

Fey, M. E., Cleave, P. L., & Long, S. H. (1997). Two models of grammar facilitation in children with language impairments: Phase 2. *Journal of Speech, Language, and Hearing Research 40,* 5-19.

Fey, M. E., Cleave, P. L., Long, S. H., & Hughes, D. L. (1993). Two approaches to the facilitation of grammar in children with language impairment: An experimental evaluation. *Journal of Speech and Hearing Research, 36,* 141-157.

Fey, M. E., Cleave, P. L., & Ravida, A. I., Long, S. H., Dejmal, A. E., & Easton, D. L. (1994). Effects of grammar facilitation on the phonological performance of children with speech and language impairments. *Journal of Speech and Hearing Research, 37,* 594-607.

Gibbard, D. (1994). Parental-based intervention with pre-school language-delayed children. *European Journal of Disorders of Communication 29,* 131-150.

Girolametto, L., Pearce, P. S., & Weitzman, E. (1996). Interactive focused stimulation for toddlers with expressive vocabulary delays. *Journal of Speech and Hearing Research, 39,* 1274-1283.

Goldstein, H. (1990). The future of language science: A plea for language intervention research. In C. Shewan (Ed.), *ASHA Reports 20: The "future of science and services" seminar* (pp. 41-50). Rockville, MD: ASHA.

Goldstein, H., English, K., Shafer, K., & Kaczmarek, L. (1997). Interaction among preschoolers with and without disabilities: Effects of across-the-day peer intervention. *Journal of Speech, Language, and Hearing Research, 40,* 33-48.

Gottschalk, M., E., Prelock, P. A., Weiler, E. M., & Sandman, D. (1997). Metapragmatic Awareness of Explanatory Adequacy II: Follow-up. *Language, Speech, and Hearing Services in Schools, 28,* 108-114.

Hart, B. (1985). Naturalistic language training techniques. In S. F. Warren & A. K. Rogers-Warren (Eds), *Teaching functional language.* (pp. 63-88) Baltimore, MD: University Park Press.

Haynes, C. (1992). Vocabulary deficit: One problem or many? *Child Language Teaching and Therapy, 8,* 1-17.

Hess, L. J., Wagner, M., DeWald, B., & Conn, P. (1993). Conversation skill intervention programme for adolescents with learning disabilities. *Child Language Teaching and Therapy, 9,* 13-31.

Hyde Wright, S. (1993). Teaching word-finding strategies to severely language-impaired children. *European Journal of Disorders of Communication, 28,* 165-175.

Hyde Wright, S., Gorrie, B., Haynes, C., & Shipman, A. (1993). What's in a name? Comparative therapy for word-finding difficulties using semantic and phonological approaches. *Child Language Teaching and Therapy, 9,* 214-229.

Iacono, T. A., Chan, J. B., & Waring, R. E. (1998). Efficacy of a parent-implemented early language intervention based on collaborative consultation. *International Journal of Language and Communication Disorders, 33,* 281-303.

Kaiser, A. P., & Hester, P. P. (1994). Generalized effects of enhanced milieu teaching. *Journal of Speech and Hearing Research, 37,* 1320-1340.

Kaufmann, S. S., Prelock, P. A., Weiler, E. M., Creaghead, N. A., & Donnelly, C. A. (1994). Metapragmatic awareness of explanation adequacy: Developing skills for academic success from a collaborative communication skills unit. *Language, Speech, and Hearing Services in Schools, 25,* 174-180.

Kim, Y. T., & Lombardino, L. J. (1991). The efficacy of script contexts in language comprehension intervention with children who have mental retardation. *Journal of Speech and Hearing Research, 34,* 845-857.

Klecan-Aker, J. S. (1993). A treatment programme for improving story-telling ability: A case study. *Child Language Teaching and Therapy, 9,* 105-115.

Klecan-Aker, J. S., Green, L. B., & Flahive, L. K. (1995). Language therapy with a child with sensory integration dysfunction: A case study. *Child Language Teaching and Therapy, 11,* 273-288.

Kluppel, D. (1991). Needed: Intervention research. In J. Miller (Ed.), *Research on child language disorders: A decade of progress*, (pp. 243-252). Austin, TX: ProEd.

Kot, A., & Law, J. (1995). Intervention with preschool children with specific language impairments: A comparison of two different approaches to treatment. *Child Language Teaching and Therapy, 11,* 144-162.

Kratochwill, T. R. (1992). Single-case research design and analysis: An overview. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 1-14). Hillsdale, NJ: Erlbaum.

Lamb, S. J., Bibby, P. A., & Wood, D. J. (1997). Promoting the communication skills of children with moderate learning difficulties. *Child Language Teaching and Therapy, 13,* 261-278.

Law, J., *Evaluating intervention for language impaired children: A review of the literature.* European Journal of Disorders of Communication, 1997. 32, (2, Special Issue): p. 1-14.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (1998). Screening for speech and language delay: a systematic review of the literature. *Health Technology Assessment, 2,* 1-179.

Layton, T. L., & Savino, M. A. (1990). Acquiring a communication system by sign and speech in a child with Down syndrome: A longitudinal investigation. *Child Language Teaching and Therapy, 6,* 59-76.

Levin, J. R. (1992). Single-case research design and analysis: Comments and concerns. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 213-224). Hillsdale, NJ: Erlbaum.

Long, S. H., Olswang, L. B., Brian, J., & Dale, P. A. (1997). Productivity of emerging word combinations in toddlers with specific expressive language impairment. *American Journal of Speech-Language Pathology, 6,* 34-47.

Medical Research Council of Canada (1998). *Tri-council policy statement: Ethical conduct for research involving humans* (Public Works and Government Services Canada No. MR21-18/1998 E). Ottawa, ON: Author.

Mann-Mandelbaum, S. E. (1990). Constraint questions: How can they be taught to children with special needs? *British Journal of Disorders of Communication, 25,* 17-33.

Masterson, J. J., & Perrey, C. D. (1999). Training analogical reasoning skills in children with language disorders. *American Journal of Speech-Language Pathology, 8,* 53-61.

McDade, A., & McCarten, P. (1998). ""Partnership with parents" A pilot project." *International Journal of Language and Communication Disorders, 33*(Suppl), 556-561.

McGregor, K. K. (1994). Use of phonological information in a word-

finding treatment for children. *Journal of Speech and Hearing Research, 37,* 1381-1393.

Nelson, K. E., Camarata, S. M., Welsh, J., & Butkovsky, L., & Camarata, M. (1996). Effects of imitative and conversational recasting treatment on the acquisition of grammar in children with specific language impairment and younger language-normal children. *Journal of Speech and Hearing Research, 39,* 850-859.

Olswang, L. B. (1998). Treatment efficacy research. In C. Frattali (Ed.), *Measuring response outcomes in speech-language pathology,* (pp. 134-150). New York, NY: Thieme.

Olswang, L. B., Thompson, C. K., Warren, S. F., & Minghetti, N. J. (1990). *Treatment efficacy research in communication disorders.* Rockville, MD: American Speech-Language-Hearing Foundation.

Roberts, J. E., Prizant, B., & McWilliam, R. A. (1995). Out-of-class versus in-class service delivery in language intervention: Effects on communication interactions with young children. *American Journal of Speech-Language Pathology, 4,* 87-93.

Robertson, S. B., & Ellis Weismer, S. (1999). Effects of treatment on linguistic and social skills in toddlers with delayed language development. *Journal of Speech, Language, and Hearing Research, 42,* 1234-1248.

Schiavetti, N., & Metz, D. E. (1997). *Evaluating research in communicative disorders* (3rd Edition). Boston, MA: Allyn & Bacon.

Schuele, C. M., Rice, M. L., & Wilcox, K. A. (1995). Redirects: A strategy to increase peer initiations. *Journal of Speech and Hearing Research, 38,* 1319-1333.

Silverman, F. H. (1998). *Research design and evaluation in speech-language pathology and audiology* (4th Edition). Needham Heights, MA: Allyn & Bacon.

Sim, I. (1998). One plus one equals three! Improving vocabulary acquisition and learning in pupils with speech and language impairments. *Child Language Teaching and Therapy, 14,* 83-93.

Sorensen, P., & Fey, M. E. (1992). Informativeness as a clinical principle: What's really new? *Language, Speech, and Hearing Services in Schools, 23,* 320-328.

Tyler, A. A., & Sandoval, K. T. (1994). Preschoolers with phonological and language disorders: Treating different linguistic domains. *Language, Speech, and Hearing Services in Schools, 25,* 215-234.

Tyler, A. A., & Watterson, K. H. (1991). Effects of phonological versus language intervention in preschoolers with both phonological and language impairment. *Child Language Teaching and Therapy, 7,* 141-160.

Venn, M. L., Wolery, M., Fleming, L. A., DeCesare, L. D., Morris, A., & Cuffs, M. S. (1993). Effects of teaching preschool peers to use the mand-model procedure during snack activities. *American Journal of Speech-Language Pathology, 2,* 38-47.

Ward, S. (1999). An investigation into the effectiveness of an early intervention method for delayed language development in young children. *International Journal of Language and Communication Disorders, 34,* 243-264.

Warren, S. F., Gazdag, G. E., Bambara, L. M., & Jones, H. A. (1994). Changes in the generativity and use of semantic relationships concurrent with milieu language intervention. *Journal of Speech and Hearing Research, 37,* 924-934.

Warren, S. F., Yoder, P. J., Gazdag, G. E., Kim, K., & Jones, H. A. (1993). Facilitating prelinguistic communication skills in young children with developmental delay. *Journal of Speech and Hearing Research, 36,* 83-97.

Weintraub, A., Lougeay-Mottinger, J., & Stillman, R. (1991). Facilitating communication development in a language-disordered child. *Child Language Teaching and Therapy, 7,* 260-272.

Wilcox, M. J., Kouri, T. A., & Caswell, S. B. (1991). Early language intervention: A comparison of classroom and individual treatment. *American Journal of Speech-Language Pathology, 1,* 49-62.

Wing, C. S. (1990). A preliminary investigation of generalization to untrained words following two treatments of children's word-finding problems. *Language, Speech, and Hearing Services in Schools, 21,* 151-156.

Wittman, S. (1996). A case study in word finding. *Child Language Teaching and Therapy, 12,*300-313.

Yoder, P. J., Kaiser, A. P., & Alpert, C. L. (1991). An exploratory study of the interaction between language teaching methods and child characteristics. *Journal of Speech and Hearing Research, 34,* 155-167.

Yoder, P. J., Warren, S. F., Kim, K, & Gazdag, G. E. (1994). Facilitating prelinguistic communication skills in young children with developmental delay: II. Systematic replication and extension. *Journal of Speech and Hearing Research, 37,* 841-851.

Yoder, P. J., Spruytenburg, H., Edwards, A., & Davies, B. (1995). Effect of verbal routine contexts and expansions on gains in the mean length of utterance in children with developmental delays. *Language, Speech, and Hearing Services in Schools, 26,* 21-32.

Yoder, P. J., & Warren, S. F. (1998). Maternal responsivity predicts the prelinguistic communication intervention that facilitates generalized intentional communication. *Journal of Speech, Language, and Hearing Research, 41,* 1207-1219.

Yoder, P. J., & Warren, S. F. (1993). Can developmentally delayed children's language development be enhanced through prelinguistic intervention? In A. P. Kaiser & D. B. Gray (Eds.), *Enhancing children's communication: Research foundations for intervention* (pp. 35-62). Balimore, MD: Paul H. Brookes.

Manuscript received: September 5, 2000
Accepted: March 3, 2001