Application of Single Subject Randomization Designs to Communicative Disorders Research

Susan Rvachew
Speech-Language Department
Alberta Children's Hospital, Alberta

Abstract
Single-subject randomization designs differ from traditional single-subject designs in that they involve random assignment of treatments to treatment times. This use of random assignment allows for the application of a randomization test to the data, thus combining the advantages of single-subject research with the advantages of statistical analysis. Hypothetical examples are used to demonstrate how four single-subject randomization designs can be applied to communicative disorders research.

Application of Single Subject Randomization Designs to Communicative Disorders Research

The value of single-subject designs for communicative disorders research has been discussed recently by McReynolds and Kearns (1983). Briefly, the primary advantages of single-subject designs over multiple-subject designs include: (1) greater control over subject variability, (2) economy, and (3) the ability to study unique patients. In addition, single-subject designs are attractive because clinicians are primarily concerned with the effects of treatments on individuals, rather than on groups of subjects. A treatment which has been shown to have a beneficial effect on group performance may not be useful with any particular patient on a clinician's caseload.

Single-Subject Randomization Designs

McReynolds and Kearns (1983) have described a variety of single-subject designs, most of which are variations on the withdrawal, reversal, or multiple baseline designs. Although these designs can be applied to a large number of clinical problems, there are many situations in which none of these designs can be used effectively. In some cases, a withdrawal of the treatment will not result in the necessary performance decrement. In many cases, the use of a reversal design will be ethically questionable. Finally, the multiple baseline (across behaviors) design cannot be used when there is only one target behavior, or when the target behaviors are very similar.

A design which can be applied in those situations where the traditional single-subject design is not feasible has been described by Edgington (1987). Single-subject randomization designs involve the random assignment of treatment times to treatments, thus controlling for variation over time in the same way that multiple-subject designs control for variation over subjects. This use of random assignment allows the researcher to apply a randomization test to the experimental results in order to determine statistical significance.

The use of statistical tests in single-subject research has been criticized on a number of grounds. Most of these criticisms are specific to the use of parametric, inferential statistical analyses, which are problematic with single-subject designs because of the serial dependence of within-subject data. Although independence of scores is a basic assumption underlying the use of parametric tests, randomization tests are valid even when the scores are not independent. Therefore, the use of randomization tests with single-subject designs involving random assignment of treatments to sessions circumvents this problem.

Other criticisms apply to the use of any statistical procedures as a substitute for visual analysis of single-subject data. It is often said that clinical researchers are concerned only with clinically significant results, and thus, statistical significance is irrelevant in applied research. Implicit in this criticism is the assumption that statistical analysis is required for the detection of small effects, but not necessary for the detection of large effects (McReynolds & Kearns, 1983). However, clinical significance and statistical significance are two quite different ways of judging experimental outcomes. Clinical significance is concerned with the size of the observed effect, while statistical significance is concerned with the source of the effect. An effect large enough to be of clinical interest may well have been caused by an extraneous variable. For this reason all experiments, including traditional single-subject experiments, require some judgement about the probability that the results reflect a true treatment effect, regardless of the size of the observed effect.

Clinical significance is determined by relating the size of the observed treatment effect to the effect size that can be expected based on one's knowledge of the clinical population under study. Often the effect will not be judged to be clinically significant unless it is fairly large. However, there are cases when small effects are of clinical interest. For example, a researcher comparing the relative effectiveness of the individual components of a treatment program may expect the program components to have relatively small effects when applied individually. Another experimenter may conclude that a small

advantage of one treatment technique over another is clinically significant when the superior treatment is less costly to apply than the inferior treatment.

Internal Validity in Multiple-Subject Designs

Clinically significant effect sizes are of no interest unless it can be shown that the observed effect is caused by the treatment. In other words, the experiment must be shown to be internally valid. In multiple-subject research, internal validity can be threatened by intersubject variability and by uncontrolled extraneous variables which are associated with the environment.

The basis for control of intersubject variability is random assignment of subjects to groups. This procedure does not, as is commonly believed, "equalize" the groups with respect to subject variables; rather, random assignment satisfies one of the primary assumptions underlying the use of statistical tests. In fact, the statistical analysis is predicated on the assumption that the groups may be different before the treatment is applied. The purpose of the statistical analysis is to determine the probability of the observed between-group difference occurring when the null hypothesis is true. When this probability is low, the researcher assumes that differences between groups are due to the treatment.

Statistical analysis does not control for extraneous variables which emanate from the environment, and consequently the researcher must take every precaution to ensure that such variables are held constant across groups or eliminated entirely. For example, when studying a treatment designed to improve reading ability, both control and treatment subjects must receive their reading test under identical lighting conditions.

Internal Validity in Traditional Single-Subject Experiments

While intersubject variability is not an issue in single-subject experiments, other sources of extraneous variation can pose a serious threat to internal validity. The first technique used to control for this kind of variation is the establishment of a stable baseline. When a stable baseline has been obtained, it is assumed that sources of extraneous variation have been brought under control, and that the subject will experience the same extraneous variables during the treatment phase as were experienced during the baseline phase.

In addition, treatment and no-treatment phases are sequenced over time in order to demonstrate experimental control of the dependent variable. If changes in the dependent variable are coincident with introduction of the treatment on more than one occasion, then the researcher can be reasonably confident that the observed effect is due to the treatment manipulation (McReynolds & Thompson, 1986).

Internal Validity in Single-Subject Randomization Designs

The designs discussed above are internally valid to the extent that the researcher can be sure that extrinsic extraneous variables have been controlled. In contrast to these designs, randomized single-subject designs do not assume that extraneous variables are constant across the treatment and control conditions. Rather, these extraneous variables are controlled in a manner similar to the control of intersubject variability in multiple-subject research. First, extraneous variables are assumed to be independent of the treatment manipulation. In other words, an extraneous variable operating during any given session is assumed to exist regardless of whether the treatment condition or the control condition is applied during that session. Secondly, the treatment and control conditions are randomly assigned to sessions, thus allowing for the application of a randomization test to the data. The randomization test yields a numerical probability value which indicates the likelihood of the obtained results occurring when the null hypothesis is true. The size of the resulting probability value determines the researcher's confidence in assuming that the observed effect was caused by the treatment.

This is not to say that a researcher using the design should ignore potential sources of extraneous variation. Extraneous variables can seriously undermine the sensitivity of the randomization test. If such sources of extraneous variation are not controlled, true treatment effects may be masked and go undetected. However, a failure to control for all possible sources of extraneous variation does not affect the internal validity of the experiment.

This paper presents hypothetical examples of single-subject experiments illustrating four different single-subject designs.[1] Although it is hoped that these examples will give the reader some idea of the potential usefulness of this design for communicative disorders research, the paper does not include sufficient detail to allow a researcher to apply any of the designs without first consulting other sources. These other sources, especially Edgington (1987), will give the reader a fuller understanding of the logic and assumptions underlying the use of randomization tests, as well as specific instructions for applying randomization tests in different experimental situations.

A Completely Randomized Design

The most basic form of this design is the completely randomized design in which any assignment of treatment times to treatments is possible (given certain sample size constraints). This design can be used to compare a treatment with no treatment, one treatment with any number of other treatments, or one component of a treatment program with other compo-

1. The purpose of these hypothetical examples is to demonstrate how the design might be used and to show how the randomization test would be applied to data resulting from similar experiments. I am not suggesting that these designs are the best way to study the problems considered in the examples. Neither am I suggesting that these hypothetical experiments could be conducted without modification.

nents of the same program. This design was used to investigate the effect of food colouring on children's behavior (Weiss, Williams, Margen, Abrams, Caan, Citron, Cox, McKibben, Ogar, & Schultz, 1980).

This example illustrates how the completely randomized design could be used to test the following hypothesis: a fluency enhancement training procedure will be more effective in reducing dysfluencies than a contingency management procedure. In this hypothetical example, the fluency enhancement (FE) treatment involves modelling and reinforcing the use of easy onset, irrespective of the presence or absence of dysfluencies in the subject's speech. The contingency management (CM) procedure involves reinforcing fluent utterances and punishing dysfluent utterances while paying no attention to speaking style. The dependent measure is the average number of dysfluencies per minute.

In order to apply the completely randomized design to this problem it is necessary to schedule a number of treatment sessions. The number and length of these sessions and the interval of time separating them will be determined by practical and theoretical considerations. In this example, six 10 minute sessions are scheduled to occur consecutively over a 1 hour period. Treatment FE is assigned at random to three of the sessions, while treatment CM is assigned to the remaining three sessions.[2] The subject is encouraged to converse on a number of topics while the experimenter applies the treatments according to the randomly determined sequence shown in Table 1. A tape recording of the six sessions is scored, yielding an average number of dysfluencies per minute for each session.

Table 1. Hypothetical outcome of a fully randomized experiment.

Session:	1	2	3	3	5	6
Treatment:	CM	FE	CM	CM	FE	FE
Score:	10	8	9	9	7	5

Note: CM = contingency management and FE = fluency enhancement (see text for a more detailed description of these treatments). The dependent variable is the mean number of dysfluencies per minute.

The null hypothesis is that the obtained scores are independent of the treatment administered at any given time. If the null hypothesis is true, any differences between scores are due to a difference in the times that the treatments were administered, and not to differences in treatment effectiveness. In order to test this hypothesis it is necessary to calculate the test statistic, which in this case, is $M_{CM} - M_{FE}$ for a one-tailed test, or the absolute value of $M_{CM} - M_{FE}$ for a two-tailed test. Since *CM* is expected to yield higher rates of dysfluency, the test statistic is $M_{CM} - M_{FE} = 2.67$.

Next, each possible way of assigning treatments to sessions is determined, and the test statistic is calculated for each of these possible assignments. In this example, there are 6!/3!3! = 20 equally probable assignments of treatments to treatment times. These assignments and the associated test statistics are called the reference set and are shown in Table 2. The probability (p) value associated with the obtained result is the proportion of test statistics greater than or equal to the obtained test statistic value[3]. Inspection of Table 2 shows that there is one test statistic equal to or greater than 2.67, and therefore, $p = 1/20 = 0.05$.

Control of Temporal Trends

The data shown in Table 1 illustrate a temporal trend: The subject's rate of dysfluency is decreasing over time in a way that does not appear to be dependent on the treatment administered at any given time. This improvement is probably due to the subject's growing familiarity with the experimenter and the experimental situation. This trend also may be due to carry over of the effects of the treatments from session to session. This kind of temporal trend may occur as a result of practice effects, fatigue effects, maturation, or historical factors.

Although this kind of consistent trend does not affect the validity of the randomization test, it may reduce the sensitivity of the statistical test. Therefore, it is important to schedule the desired number of treatment sessions in a way that will minimize the likelihood of such a trend occurring. With respect to the previous example, the treatment sessions could have been scheduled to occur once a week over a six week period. In addition, a different examiner could have been assigned to each of the sessions.

2. A researcher may wish to observe the effects of each treatment as they occur over time. In this case, each "session" would consist of a longer period of time (e.g., a week or a month) during which repeated applications of the treatment (or control condition) and repeated probes would occur. A graphic representation of the data would then show multiple data points per session. For statistical analysis, however, all of the data collected during each session would be treated like a single data point.

3. Although this statistical test can be calculated by hand, randomization tests are usually performed by computer. The computer programs necessary for performing randomization tests can be found in Edgington (1987). The specific programs used to determine the p values for the first three examples were Program 4.4 for the completely randomized design, Program 4.2 for the randomized block design, and Program 6.1 for the factorial design. These programs were developed for use with multiple-subject designs but are equally appropriate for use with single-subject data. A computer program, written in Basic, for calculating the test used in the last example can be obtained from the author.

Table 2. Reference set for example 1.

Session:	1	2	3	4	5	6	Test Statistic
Score:	10	8	9	9	7	5	
1.	CM	CM	CM	FE	FE	FE	2.000
2.	CM	CM	FE	CM	FE	FE	2.000
3.	CM	CM	FE	FE	CM	FE	.667
4.	CM	CM	FE	FE	FE	CM	-.667
5.	CM	FE	CM	CM	FE	FE	2.667*
6.	CM	FE	CM	FE	CM	FE	1.330
7.	CM	FE	CM	FE	FE	CM	0.000
8.	CM	FE	FE	CM	CM	FE	1.330
9.	CM	FE	FE	CM	FE	CM	0.000
10.	CM	FE	FE	FE	CM	CM	-1.330
11.	FE	CM	CM	CM	FE	FE	1.330
12.	FE	CM	CM	FE	CM	FE	0.000
13.	FE	CM	CM	FE	FE	CM	-1.330
14.	FE	CM	FE	CM	CM	FE	0.000
15.	FE	CM	FE	CM	FE	CM	-1.330
16.	FE	CM	FE	FE	CM	CM	-2.667
17.	FE	FE	CM	CM	CM	FE	.667
18.	FE	FE	CM	CM	FE	CM	-.667
19.	FE	FE	CM	FE	CM	CM	-2.000
20.	FE	FE	FE	CM	CM	CM	-2.000

*This line shows the actual assignment of treatments to sessions and the obtained test statistic value.

Randomized Block Design

An alternative to the completely randomized design which provides greater control of systematic variation over time is the randomized block design. This design involves dividing the total number of treatment sessions into blocks, and then administering all of the treatments within each block. The order in which the treatments are administered is randomly determined for each individual block. Smith (1963) used this design (with a different statistical analysis) to study the effects of three drugs on narcolepsy. This example illustrates how this design could be used to study the relative effects of different language stimulation techniques.

In the following hypothetical example, three different language stimulation techniques will be used with a young, language-delayed child: (1) providing social reinforcement contingent upon correct imitations (IM) of single-words in the presence of an appropriate model and stimulus object; (2) providing parallel talk (PT) at the single-word level while observing the child at play with a variety of objects; and (3) playing turn-taking (TT) games with the child in which the examiner follows the child's lead while providing appropriate verbal models when taking his or her turn. The dependent measure is a count of the number of different words spoken by the child during each session.

Treatment sessions are scheduled for each Monday, Wednesday, and Friday morning over a 4 week period. Each week of sessions constitutes a block, and each of the 3 days is assigned, at random, to one of the three treatments. This random assignment of days to treatments is made individually for each block, yielding the sequence of treatments shown in Table 3. The number of new words spoken during each session also is shown. These data clearly demonstrate a temporal trend which could be due to maturation on the part of the child and the cumulative effects of the treatments over time. However, the particular random assignment used assures that the effects of this trend are more or less equal for each of the three treatments and allows for statistical control of this trend.

This design is analogous to a repeated measures design in which four subjects receive each of three treatments, and therefore, the appropriate test statistic is F for repeated measures analysis of variance. For this example, $F = 4.69$. There are $(3!)^4 = 1{,}296$ equally probable assignments, 36 of which yield a test statistic greater than or equal to 4.69; therefore, $p = 0.03$.

Factorial Design

The development of phonological process treatment approaches has led to a resurgence of interest in minimal pair contrasting and auditory bombardment as techniques for the remediation of speech sound errors (Tyler, Edwards, & Saxman, 1987). The single-subject randomization design could be used to study both of these treatment variables with the same subject. In this hypothetical example, a subject who presents with fricative stopping is scheduled to participate in two half-hour treatment sessions per week over a 6 week period. Each session consists of the following activities: (1) a pre-treatment probe which measures the proportion of stopped fricatives observed during a picture discussion task; (2) auditory bombardment with stimuli which are either appropriate (Auditory Bombardment - Treatment; ABT) or inappropriate (Auditory Bombardment - Control; ABC), given the target process; (3) phonological process therapy using a minimal pairs technique (Therapy - Treatment; THT), or articulation therapy using traditional techniques (Therapy - Control; THC); and finally, (4) a post-treatment probe.

Six of the treatment sessions are randomly assigned to the ABT condition while the remaining sessions are assigned to the

Table 3. Hypothetical outcome of a randomized block experiment.

Block (Week):	Block 1			Block 2			Block 3			Block 4		
Day:	M	W	F	M	W	F	M	W	F	M	W	F
Treatment:	IM	PT	TT	PT	IM	TT	IM	TT	PT	TT	IM	PT
Score:	1	3	7	5	6	8	7	10	9	12	10	12

Note: M = Monday, W = Wednesday, F = Friday, IM = imitation with reinforcement, PT = parallel talk, TT = turn-taking games (see text for a more detailed description of these treatments). The dependent variable is the number of different words spoken by the child during the session.

ABC procedure. Next, three of the ABT sessions are randomly assigned to THT, and the remaining three ABT sessions are assigned to the THC procedure. This process is then repeated with the six ABC sessions. One possible outcome of this random assignment procedure is shown in Table 4. Table 4 also shows the hypothetical results of this experiment.

it is expected that the traditional procedure will provide the larger measurements, indicating less improvement in phonological ability (see Edgington (1987) for a more detailed discussion of test statistics for factorial randomization designs). The number of test statistics in the reference set is the number of possible assignments for the ABT condition, times the number

Table 4. Hypothetical outcome of a 2 x 2 factorial experiment.

Session:	1	2	3	4	5	6	7	8	9	10	11	12
Variable 1:	ABC	ABT	ABT	ABC	ABC	ABC	ABT	ABT	ABT	ABT	ABC	ABC
Variable 2:	THC	THT	THT	THT	THC	THT	THT	THC	THC	THC	THT	THC
Pre-test:	.10	.10	.10	.20	.25	.30	.30	.40	.45	.50	.50	.60
Post-test:	.30	.20	.20	.50	.40	.50	.40	.55	.60	.70	.90	1.00
Difference Score:	.20	.10	.10	.30	.15	.20	.10	.15	.15	.20	.40	.40

Note: ABT = auditory bombardment - treatment (auditory bomardment with stimuli that contain the target phoneme), ABC = auditory bombardment - control (auditory bombardment with stimuli that do not contain the target phoneme), THT = therapy - treatment (minimal pairs training), THC = therapy - control (traditional articulation therapy). The dependent variable is 1 - the proportion of stopped fricatives in a speech sample.

Since a consistent improvement in probe scores can be expected to occur over the course of this experiment, the dependent measure is the difference between the post- and pre-treatment probe scores for each session. Table 5 shows these difference scores rearranged into cells representing the four possible combinations of treatment conditions.

As with a traditional factorial design, we can now test for the main effect of each of the variables over all levels of the other variable. For example, we could compare the effects of the speech therapy procedures over both types of auditory bombardment. For a one-tailed test, the test statistic is the total of the measurements in the second row of the table, or 1.25, because

of possible assignments for the ABC condition (i.e., 6!/3!3! x 6!/3!3! = 20 x 20 = 400). For this example, the null hypothesis is accepted because $p = 0.32$, one-tailed.

The statistical test used for the completely randomized design can be applied to these data to compare levels of either variable within any level of the other variable. For example, in order to examine the effect of the auditory bombardment variable for the minimal pairs training sessions, the test described for the first example above would be applied to the six measurements shown in the top row of Table 5. In this case the test statistic is 0.27 - 0.10 = 0.17, because the ABC condition is expected to provide larger measurements than the ABT condi-

Table 5. Hypothetical factorial experimental data arranged for statistical analysis.

	Auditory Bombardment	
	Control (ABC)	Treatment (ABT)
Minimal Pairs Training	0.20	0.10
(THT)	0.20	0.10
	0.40	0.10
	$M = 0.27$	$M = 0.10$
Traditional Therapy	0.20	0.15
(THC)	0.15	0.15
	0.40	0.20
	$M = 0.25$	$M = 0.17$

tion. There are 6!/3!3! = 20 equally probably assignments, and only one of these yields a test statistic value equal to or greater than 0.17. Therefore, $p = 1/20 = 0.05$, one-tailed, and the null hypothesis is rejected.

Test for Treatment Intervention

This design is superficially similar to the non-experimental AB single-subject design: Initially, the subject's performance on the dependant measure is monitored over a period of time during which no treatment is administered; then the treatment is introduced and maintained over a number of sessions, while monitoring of the subject's performance on the dependent measure continues. The subject's performance during treatment sessions is compared with his or her performance during control sessions.

Usually, the treatment is introduced when the subject's performance during control sessions has stabilized. This procedure does not allow for statistical analysis of the data because the assignment of treatment sessions to the control or treatment condition is non-random. However, if the decision to introduce the treatment is made by randomly selecting the session during which this will occur, a randomization test can be applied to the data, and the design becomes fully experimental.

This design is especially useful when the treatment under study is expected to produce permanent, or relatively permanent, effects which would carry over to control sessions. For this reason, this design was chosen to test the hypothesis that a sound identification training procedure would facilitate sound production learning by children with functional articulation errors (specifically, the substitution of /s/ for /ʃ/ or the substitution of /θ/ for /s/). Although this experiment has been carried out and statistically significant results were obtained, the following example uses hypothetical data in order to simplify the demonstration.[4]

Table 6. Hypothetical outcome of a test for treatment intervention.

Session:	1	2	3	4	5	6	7	8	9	10
Score:	10	10	40	20	10	10	30	10	20	30
Session:	11	12	13	14	15	16	17	18	19	20
Score:	50	50	50	50	50	50	40	60	40	60
Session:	21	22	23	24	25	26	27	28	29	30
Score:	60	60	60	60	70	80	80	70	60	50

Note: Underlined scores were obtained during treatment sessions, while the other scores were obtained during control sessions.

Each treatment session lasted no longer than ten minutes and consisted of repeated sound identification training trials (using either control or treatment stimuli) followed by a 10 item, imitative production probe. Three 1 hour periods, scheduled to occur on consecutive days, were required to complete 30 such treatment sessions. The first five sessions were reserved for the control condition, and the last five sessions were reserved for the treatment condition. The starting session for treatment intervention was chosen randomly from among sessions 6 to 25.

The hypothetical results of this experiment are shown in Table 6. For this example, the control procedure was administered during sessions 1 through 10, and the treatment procedure was administered during sessions 11 through 30. The number of correct productions of the target phoneme during the production probe are shown for each session. The measures obtained during treatment sessions are underlined.

The one-tailed test statistic is the difference between the mean probe score for treatment sessions and the mean probe score for control sessions, which in this case is 5.75 - 1.90 = 3.85. This test statistic was calculated for each of the 20 possible assignments of sessions to treatments. Only one test statistic in the reference set is equal to or greater than 3.85, and consequently $p = 1/20 = 0.05$, one-tailed.

4. A detailed description of this study is currently being prepared for publication. More information about the sound identification training procedure and the actual results of the study can be obtained from Dr. D. G. Jamieson, Speech Communication Laboratory, Department of Communicative Disorders, Elborn College, University of Western Ontario, London, Ontario, N6G 1H1, or from the author.

Conclusion

Recently the difficulties inherent in the application of basic, traditional single-subject designs have been discussed, and the need for flexibility in the use of single-subject designs has been noted (Connell & Thompson, 1986; Kearns, 1986). The single-subject randomization design adds to the flexibility of single-subject designs. It can be used in any situation where the traditional single-subject design might be applied. It also can be used in situations where the traditional designs would be difficult to apply (e.g., a factorial single-subject experiment or an experiment in which the treatment has irreversible effects). The examples discussed above are only some of many possible randomized single-subject designs. In fact, a randomization test can be applied to any conceivable single-subject experiment in which there is random assignment of treatment times to treatments. Further information and additional examples can be found in Edgington (1984), Edgington (1987), Kazden (1976), and Levin, Marascuito, and Hubert (1978).

Address all correspondence to:
Susan Rvachew
Speech-Language Department, Alberta Children's Hospital
1820 Richmond Rd. S.W.
Calgary, Alberta, T2T 5C7

Acknowledgements

I am grateful to Dr. Edgington (Department of Psychology, University of Calgary) for his comments on this manuscript.

References

Connell, P. J. & Thompson, C. K. (1986). Flexibility of single-subject experimental designs. Part III: Using flexibility to design or modify experiments. *Journal of Speech and Hearing Disorders, 51*, 214-225.

Edgington, E. S. (1984). Statistics and single case analysis. In M. Hersen, R. M. Eisler, & P. Miller (Eds.). *Progress in Behavior Modification*, (pp. 83-119, Vol.16). New York: Academic Press, Inc..

Edgington, E. S. (1987). *Randomization Tests* (2nd edition). New York: Marcel Dekker, Inc.

Kazdin, A. E. (1976). Statistical analyses for single-case experimental designs. In M. Hersen & D. H. Barlow (Eds.). *Single-case Experimental Designs: Strategies for Studying Behavior Change*. Oxford: Pergamon.

Kearns, K. P. (1986). Flexibility of single-subject experimental designs. Part II: Design selection and arrangement of experimental phases. *Journal of Speech and Hearing Disorders, 51*, 204-214.

Levin, J. R., Marascuito, L. A., & Hubert, L. J. (1978) Nonparametric randomization tests. In T. R. Kratochwill, (Ed.), *Single Subject Research: Strategies for Evaluating Change*. New York: Academic Press.

McReynolds, L. V., & Kearns, K. (1983). *Single-subject experimental designs in communicative disorders*. Austin, Texas: Pro-Ed.

McReynolds, L. V. & Thompson, C. K. (1986). Flexibility of single-subject experimental designs. Part I: Review of the basics of single-subject designs. *Journal of Speech and Hearing Disorders, 51*, 194-203.

Smith, C. M. (1963). Controlled observations on the single case. *Canadian Medical Association Journal, 88*, 410-412.

Tyler, A. A., Edwards, M. L., & Saxman, J. H. (1987). Clinical applications of two phonologically based treatment procedures. *Journal of Speech and Hearing Disorders, 52*, 393-409.

Weiss, B., Williams, J. H., Margen, S., Abrams, B., Caan, B., Citron, L. J., Cox, C., McKibben, J., Ogar, D., & Schultz, S. (1980). Behavioral responses to artificial food colors. *Science, 297*, 1487-1489.