Learning What Works in Sensory Disabilities: Establishing Causal Inference

John B. Cooney, John Young III, John L. Luckner, and Kay Alicyn Ferrell

Abstract: This article is intended to assist teachers and researchers in designing studies that examine the efficacy of a particular intervention or strategy with students with sensory disabilities. Ten research designs that can establish causal inference (the ability to attribute any effects to the intervention) with and without randomization are discussed.

Since the passage of the No Child Left Behind Act of 2001 (2002) and the subsequent reauthorization of the Individuals with Disabilities Education Act (2004), the field of special education has placed increasing emphasis on scientifically based research and evidence-based practice. The leading journal in special education, *Exceptional Children*, devoted an entire issue to specifying quality indicators for various research designs (Brantlinger, 2005; Gersten, 2005; Horner, 2005; Odom, 2005; Thompson, 2005), as a guide for evaluating research studies and improving the evidence base in special education.

In sensory disabilities, these quality indicators are often difficult to achieve. Because sensory disabilities are a low-prevalence disability—in 2012 to 2013, students with visual impairments comprised 0.1% of the school-age population (ages 3 to 21), while children who were deaf and hard of hearing comprised 0.2% (Snyder & Dillow, 2015)—conducting research is made more difficult not only by the small numbers, but by the fact that those numbers are dispersed over a huge geographic area. Many school districts in rural areas enroll no students with sensory disabilities; others may enroll only one student who has only one form of sensory disability. Accessing these students for the purpose of research incurs considerable expense. In addition, children with sensory disabilities are considered a heterogeneous population, meaning that two children of the same age, diagnosis, and grade level may have differing degrees of visual impairment or hearing loss that affect their ability to respond to any particular intervention. The low-prevalence nature of sensory disabilities means that replication studies are few and far between, since the same children cannot be used for replication, even with different investigators, and because the passage of time leads to confounding variables of prior exposure and maturation.

A number of articles have examined the level of evidence for various research studies in sensory disabilities in order to

identify the much-desired evidence-based practices. These include Abou-Gareeb, Lewallen, Bassett, & Courtright, 2001; Botsford (2013); Cawthon and Leppo (2013); Ferrell, Buettel, Sebald, and Pearson (2006); Ferrell, Dozier, and Monson (2011); Ferrell, Mason, Young, and Cooney (2006); Graeme et al. (2011); Kelly and Smith (2011); Luckner (2006); Luckner and Cooke (2010); Luckner and Handley (2008); Luckner, Sebald, Cooney, Young, and Muir (2005, 2006); Luckner and Urbach (2012); Parker, Davidson, and Banda (2007); Parker, Grimmett, and Summers (2008); Parker and Ivy (2014); Parker and Pogrund (2009); Wang and Williams (2014); and Wright, Harris, and Sticken (2010). All of these studies have concluded that pedagogy in sensory disabilities is characterized by a dearth of scientific evidence. A recent analysis conducted for the Collaboration for Effective Educator Development, Accountability, and Reform (CEE-DAR) project at the University of Florida, which assists states and institutes of higher education to reform their teacher preparation programs, found that most of the research in visual impairment, deaf education, and deafblindness demonstrated low levels of evidence (Ferrell, Bruce, & Luckner, 2014). There were some areas where evidence could be considered emerging, such as the body of literature that questions a school district's practice of employing para-educators (often referred to as paraprofessionals) to serve students with visual impairments (Conroy, 2007; Forster & Holbrook, 2005; Griffin-Shirley & Matlock, 2004; Harris, 2011; Koenig & Holbrook, 2000; Lewis & McKenzie, 2010; Marks, Schrader, & Levine, 1999; McKenzie & Lewis, 2008; Russotti & Shaw, 2001), because it was composed of limited intervention studies and several opinion pieces. There were also areas that were considered strong, such as early identification and intervention services prior to six months of age for infants who are deaf or hard of hearing (Joint Committee on Infant Hearing, 2007; Meinzen-Derr, Wiley, & Choo, 2011; Moeller, 2000; Vlastarakos, Proikas, Papacharalampous, Exadaktylou, Mochloulis, & Nikolopoulos, 2010; Vohr et al., 2008; Yoshinaga-Itano, 2003a, 2003b; Yoshinaga-Itano & Gravel, 2001; Yoshinaga-Itano, Sedey, Coulter, & Mehl, 1998), which were supported by a number of studies that were interventions and replicated by other researchers. Although the research base appears to be growing, Ferrell et al. (2014) concluded that research in sensory disabilities was generally characterized by studies that lacked causal inference and provided little evidence for the strategies that constitute practice.

In addition to the dearth of evidence-based claims about the effectiveness of interventions with persons with sensory disabilities, research on interventions derived from theoretical models is also noticeably absent. Moreover, there were few, if any, replications or systematic investigations of any particular class of interventions designed to enhance the education of students with sensory disabilities. This lack of such replications or investigations does not mean there is not a rich body of knowledge about these persons; however, there is relatively little scientific evidence concerning the efficacy of interventions that promote learning and development in children with sensory disabilities.

This article describes strategies for designing investigations that strengthen causal claims about interventions in-

tended for individuals with sensory disabilities. First, we describe tactics for implementing what is considered by many to be the strongest design for drawing causal inferences about interventions: a design that uses a pretest, posttest, and random assignment of study participants to experimental and control groups. Because this particular design, although considered the "gold standard" in educational research (Sullivan, 2011), is not always possible with sensory disability populations, and it is not even particularly relevant (Cronbach, 1982), we next describe some enhancements to quasi-experimental designs where the participants are not randomly assigned to experimental and control groups, but which strengthen causal inferences. The regression discontinuity design, for example, has much to offer researchers who work with individuals with sensory disabilities; however, the Ferrell (2006) and Luckner (2006) analyses did not identify a single study using this design. Hence, a third objective is to introduce the regression discontinuity design as a useful technique for constructing causal claims about intervention strategies. Fourth, we explore the use of two designs in which participants serve as their own control group: the single-factor within-subject design and the time-series design. Although single-case designs were not among the designs identified by Valentine and Cooper (2004) as meeting evidence standards for making causal claims at the time of the Ferrell and Luckner reviews, current evidence standards (Valentine & Cooper, 2008; What Works Clearinghouse, 2014) offer guidelines for constructing causal claims from single-case designs. Nevertheless, there are other research designs that can be used to demonstrate the efficacy of an intervention.

Questions in research methodology

It has become all too fashionable to frame debates in education around research methodology. Instead, we believe, as do Bartlett et al. (2006), that it is more productive to think about the selection of a research methodology in relation to research questions and the interrelationships among the questions. Shavelson, Towne, and the Committee on Scientific Principles for Education Research (2002), for example, suggested that research in education, social sciences, and natural sciences takes one of three forms: (1) understanding what is happening; (2) detecting a causal relationship; and (3) explaining a causal relationship. Viewed this way, it makes little sense to quibble about whether one should use qualitative or quantitative methodology until one has given careful thought to the questions one seeks to answer.

UNDERSTANDING WHAT IS HAPPENING

To understand what is happening, researchers need to use the tools that produce rich descriptions of persons, their environments, and their interactions, as well as those that provide accurate quantitative estimates of important characteristics. On the one hand, researchers might seek to understand the experiences of persons with sensory disabilities in segregated educational settings and how they are different from those in inclusive settings. Descriptions emerging from interviews and ethnographic studies would be

essential in accumulating information related to answering this question. On the other hand, investigators might seek to describe the characteristics of the population being served in the different settings based on gender, ethnicity, age, severity of disability, literacy level, and so on. Accurate descriptions of this nature require appropriate sampling methods, as well as descriptive and inferential statistical methods. Methods that estimate the magnitude of relationships among measured or inferred characteristics may also provide information that is useful for understanding what is happening.

DETECTING A CAUSAL RELATIONSHIP

When the focus of research questions turns to ascertaining causal effects, experiments that use random assignment to form treatment and control groups provide the most logically defensible no-cause baseline against which to estimate the effects of an intervention (Cook, 2002). Although experimental designs that use random assignment provide this baseline, it is important to acknowledge there are instances in which randomized experiments are simply not possible or may be premature. It may be important to know, for example, whether the magnitude of treatment effects varies with the severity of the sensory impairment. Sensory impairment, however, is a preexisting condition that is not amenable to random assignment. One must rely on analytical (for instance, structural equation modeling) and logical methods for making causal claims (see Thompson, Diamond, McWilliam, Snyder, & Snyder, 2005, for further discussion).

Finally, it should be noted that descriptive studies that address the question of what is happening also play a key role in establishing the internal validity of experimental research. Descriptive research addresses the issues of what was actually happening in the implementation of the intervention. Questions related to the fidelity of the implementation of an intervention in experimental research play a key role in answering questions that seek to explain causal relationships.

EXPLAINING A CAUSAL RELATIONSHIP

When consistent evidence of a causal relationship is found in randomized experiments, the focus shifts to questions that seek to explain the causal process. If there are differences in literacy development in students educated in inclusive settings relative to self-contained settings, researchers want to know why that is so. In this instance, descriptive research plays a critical role in generating potential explanations for hypothesized causal relationships. Descriptions of the interactions among people in self-contained and inclusive settings, or studies of the way in which cognitive processes develop in the two settings through studies that enable learners to externalize their cognitive processes as they attempt to read and write, are likely to provide insights into strategies that are key processes leading to the development of literacy. Once the cognitive processes are clearly defined through descriptive studies, interventions may be designed around them and their effectiveness evaluated in controlled experiments.

OVERCOMING BARRIERS TO CAUSAL INFERENCE

Designing and implementing the simplest of research designs that support causal inference is a daunting task.
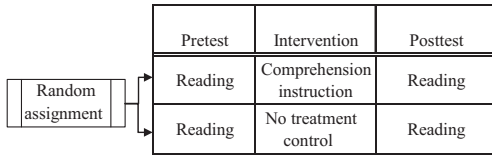
	Pretest	Intervention	Posttest
Random assignment	Reading	Comprehension instruction	Reading
	Reading	No treatment control	Reading

Figure 1. Randomized experiment with pretest and posttest.

Thus, the purpose of an investigator is not to find fault with the published research, but to encourage continuing refinements in the research that is conducted to find ways that provide incontrovertible evidence of efficacy. There are two basic research designs that meet evidence standards for causal validity: randomized experiments and regression discontinuity designs. Quasi-experiments with equivalent groups meet evidence standards with reservations. [The design standards for both regression discontinuity designs and single-case designs are still under review and are considered pilots (What Works Clearinghouse, 2014).] Quasi-experimental designs that can establish causal validity include nonequivalent groups with multiple pretests, single-factor within-subject designs, and single-case designs, all described below.

Randomized experiments

Components of the randomized experiments that support claims about causal validity are shown in Figure 1. The most crucial component of this research design is the random assignment of study participants to the treatment and no-treatment control conditions. Random assignment of participants to treatment and control groups is the best method for equalizing any differences between two groups of participants at the outset of the study.

Thus, the group that does not receive the intervention serves as the most logically defensible no-cause baseline for evaluating the effects of the intervention. Although it is not necessary to include a pretest, including one can be used to increase the power of the statistical test to detect the presence of treatment effects when treated as a covariate by reducing error variance.

The What Works Clearinghouse (2014) design standards require a minimum of 350 individuals for a randomized controlled trial, which is largely unattainable in sensory disability fields. However, an obvious objection to randomized experiments is that a potentially beneficial intervention needs to be withheld from some participants. Even though the efficacy of the intervention is not known and in fact could be detrimental, and even though a lottery is the fairest way of making the assignment, the use of such designs is questioned on ethical grounds (Cook & Campbell, 1979). Indeed, there are many other philosophical and practical objections to the use of randomized experiments (Cook, 2002). Because quasi-experiments involve withholding an intervention, and because they do not provide strong evidence for causal validity, it is difficult to justify choosing quasi-experiments over randomized experiments.

One strategy for overcoming objections to withholding an intervention that is perceived as valuable is to use a wait-list design that simply delays the introduction of the intervention for one group of participants. An example of this design is shown in Figure 2. The design uses random assignment to place study participants into
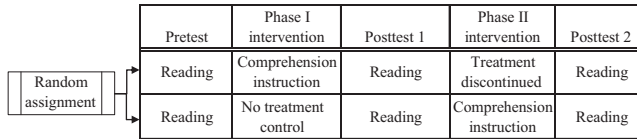
	Pretest	Phase I intervention	Posttest 1	Phase II intervention	Posttest 2
Random assignment	Reading	Comprehension instruction	Reading	Treatment discontinued	Reading
	Reading	No treatment control	Reading	Comprehension instruction	Reading

Figure 2. Randomized wait-list control group experiment with pretest and multiple posttests.

either the immediate (Phase I) or delayed (Phase II) intervention conditions and uses multiple posttests.

Another strategy for overcoming objections to withholding treatment is to actually conduct two studies within a study. Atkinson (1968) randomly assigned students to receive computer-assisted instruction either in reading or in mathematics. Both groups completed pretests and posttests in both subjects and, later, the interventions were switched. The group that was receiving computer-assisted instruction in reading was switched to receive computer-assisted instruction in mathematics, whereas the students originally receiving computer-assisted instruction in mathematics were switched to receive computer-assisted instruction in reading. After receiving the second treatment in Phase II, both groups completed a second posttest in both subjects. Figure 3 depicts this multiple treatment design. In this design, each group serves as a control (the no-cause baseline) for the other group in order to estimate the impact of the interventions. The key to using this design is that the two interventions are independent in terms of their impact on performance in the domains being assessed. (Although students' reading and mathematics achievement are likely to be correlated, it is not likely that instruction in mathematics will affect reading, or that reading instruction will impact mathematical reasoning.) Insofar as both interventions are perceived as valuable, the design overcomes objections to withholding treatment even when only the first phase is completed. Precision in the estimation of the magnitude of the effect of treatment can be increased in the analysis by using students' pretest scores in reading and mathematics as covariates in an analysis of covariance model.

Randomized experiments with planned variation designs

A variant of the multiple treatment design is the planned variation design (Rivlin & Timpane, 1975), although it, too, is subject to the minimum 350 participants by design standards (What Works Clearinghouse, 2014). Continuing with the example above, an investigator may vary the level or amount of the intervention strategy. For example, the investigator may fix the amount of instructional time to 20 minutes on one, three, or five days a
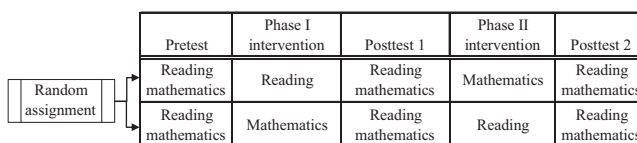
	Pretest	Phase I intervention	Posttest 1	Phase II intervention	Posttest 2
Random assignment	Reading mathematics	Reading	Reading mathematics	Mathematics	Reading mathematics
	Reading mathematics	Mathematics	Reading mathematics	Reading	Reading mathematics

Figure 3. Randomized experiment with multiple treatments, pretest, and multiple posttests.

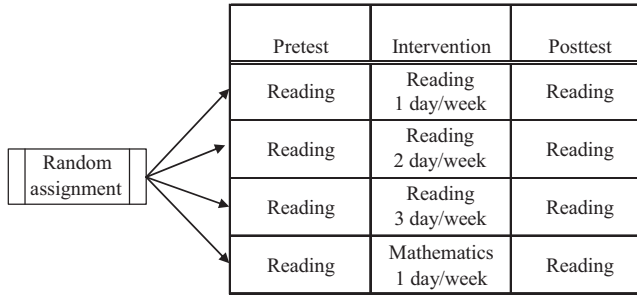|	Pretest	Intervention	Posttest
	Reading	Reading 1 day/week	Reading
Random assignment	Reading	Reading 2 day/week	Reading
	Reading	Reading 3 day/week	Reading
	Reading	Mathematics 1 day/week	Reading

Figure 4. Randomized experiment with planned variation of intervention.

week (an approach reminiscent of the dose-response design often used in medical research). In the absence of a control group, however, this design only gives an estimate of the relative impact of an intervention. Adding a control group that receives a valued intervention that does not impact the outcome of interest provides the necessary no-cause baseline to evaluate the absolute effect (Cook & Campbell, 1979). The planned variation design is depicted in Figure 4. Notice that a control group receiving an intervention for mathematics is included. Insofar as the mathematics intervention is not expected to impact the reading outcome measure, it may be considered a no-cause baseline to judge the effects of varying levels of treatment.

Quasi-experiments

Nonequivalent group designs. Research designs based on intact groups of participants to form experimental and control groups were by far the most frequently

used designs for research on literacy among individuals with sensory disabilities. Figure 5 depicts the nonequivalent group design that is encountered most often. The design usually includes an untreated control group and pretest-posttest measures. Because the process of forming the groups is unknown and participants cannot be matched on every characteristic that may affect outcomes, there are many potential differences between the two groups that could appear as intervention effects even when investigators take extraordinary steps to match the two groups and randomly assign the groups to experimental and control conditions. Depending on the outcome of the experiment, the results may or may not be interpretable (Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002). The four major threats to causal inferences that one must eliminate in this design include: selection-maturation, instrumentation, differential statistical regression, and local history.
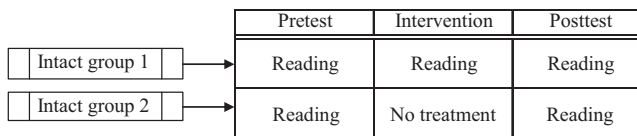
	Pretest	Intervention	Posttest
Intact group 1	Reading	Reading	Reading
Intact group 2	Reading	No treatment	Reading

Figure 5. Nonequivalent groups—quasi-experiments with pretests and posttests—that are most often encountered.
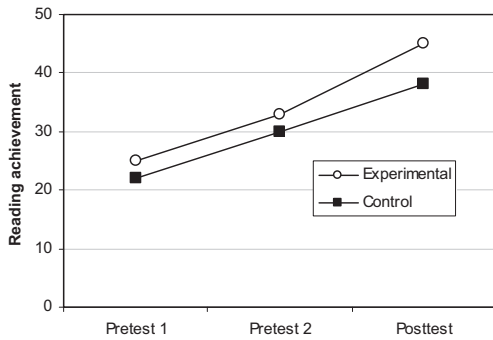
Figure 6. Hypothetical results of an experiment with multiple pretests.

Nonequivalent groups with multiple pretests. The interpretation of nonequivalent designs with a control group can be enhanced considerably by adding one or more pretests to the design, as shown in Figure 6. Including an additional pretest can be useful for eliminating threats to causal inferences based on the factors of selection and maturation and statistical regression. If the two groups exhibit similar levels of growth between the two pretests, this would eliminate the rival hypothesis of differential maturation between the two groups. Consider the hypothetical results in Figure 7 compared to the design in Figure 6. If the second pretest was the only pretest, one could argue that the experimental group was already at an advantage and that its superior performance on the posttest reflected the earlier advantage and had nothing to do with the intervention. The additional pretest, however, leads to a different interpretation. Although the experimental group exhibits a slight advantage in reading achievement on the first pretest, both the experimental and control groups are developing at approximately the same rates between the two pretests. Thus, differential maturation can be ruled out as an explanation for the differences between the two groups on the posttest, although it is necessary to account for the differences between the two groups on the pretest measures as well as their correlation.

Regression discontinuity designs. Given the simplicity of this design and its ability to control threats to internal validity, it is surprising that it is not used more often (the Ferrell, 2006, and Luckner, 2006, studies did not identify one study using a regression discontinuity design). Suppose that an individual has administered a reading assessment at the beginning and end of the school term. For a skill like reading, it would not be unreasonable to observe a moderate to high positive correlation between the two sets of scores like that exhibited in Figure 8. Further, suppose that a researcher decides to assign all students with reading scores below a certain cutoff (the mean, for example), to receive a literacy intervention. If the intervention were effective, how would it affect the expected relationship between pretest and posttest scores? Figure 9 shows what a simple main effect of intervention would look like. If the treatment were effective, there would be a discontinuity in the regression line describing the relationship between
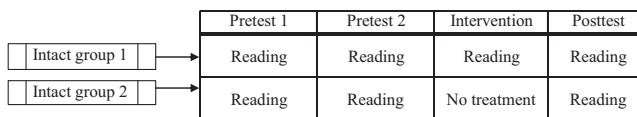
	Pretest 1	Pretest 2	Intervention	Posttest
Intact group 1	Reading	Reading	Reading	Reading
Intact group 2	Reading	Reading	No treatment	Reading

Figure 7. Nonequivalent groups, quasi-experiment, with multiple pretests and posttest.
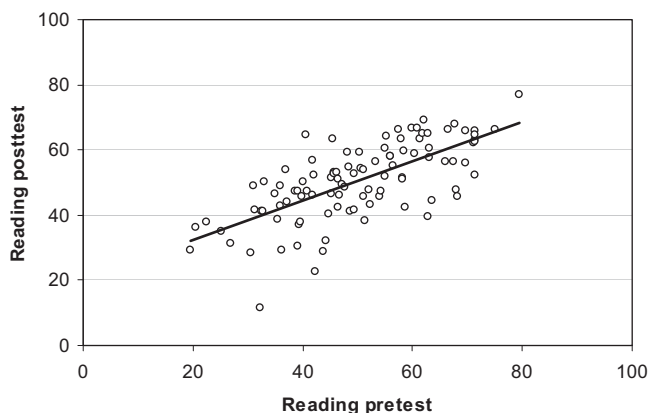
Figure 8. Hypothetical effects of intervention on the relationship between students' pretest and posttest reading assessment scores.

pretest and posttest scores. Lord and Novick (1968) show that the discontinuity in the regression line is an unbiased estimate of the intervention effect if the basic requirements of the design can be satisfied:

1. If, and only if, the assignment of participants is based on the cut-off score;
2. The cut-off score is the mean of the distribution; and
3. The functional form of the relationship between the assignment variable and the dependent variable is known.
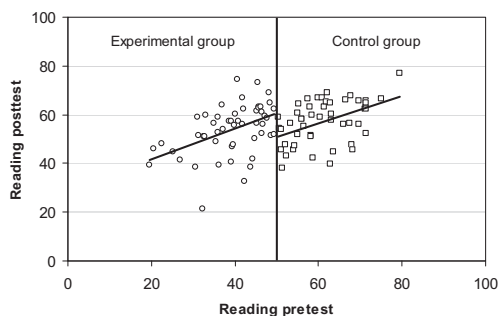

Figure 9. Hypothetical effects of intervention on the relationship between students' pretest and posttest reading assessment scores.

More information on regression discontinuity designs can be obtained from the Institute of Education Sciences' online *Technical Methods Report* (2008).

Single-factor within-subject designs. In some cases, evidence for causal validity can be demonstrated using a design where participants serve as their own control group by participating in all conditions of the experiment (Keppel & Zedeck, 1989). In Figure 10, participants are randomly assigned to receive both interventions, but in two different orders. Wauters (2001), for example, investigated the question of whether deaf students (mean hearing loss 104 dB) recognized written words learned through sign (Dutch sign language) and spoken language more effectively than when written words were learned solely through speech. Participants were trained in both conditions in different orders and with different word lists. Both the order of interventions and word lists were counterbalanced to control for practice effects and word list effects. In the absence of counterbalancing, any differences between the two
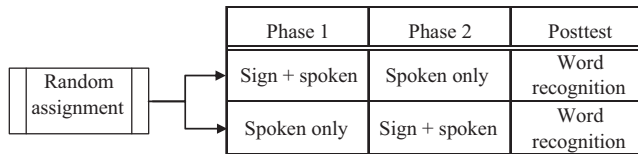
Phase 1	Phase 2	Posttest
Sign + spoken	Spoken only	Word recognition
Spoken only	Sign + spoken	Word recognition

Figure 10. Randomized within-subjects experiment.

learning conditions could be attributed to their order or to some unknown differences between the word lists. Wauters (2001) went one step further to include a third word list that was not part of training. Thus, the two learning conditions could be evaluated in absolute terms as well as in relative terms.

There is clearly one condition, however, where the single-factor within-subject design would not be considered appropriate. When an intervention has a lingering or carryover effect, it is possible that it could interact with one of the other conditions to produce a differential carry-over effect. In this case, counterbalancing does not control this threat to causal validity. For example, suppose a researcher is investigating the effects of a reading comprehension strategy (A) relative to the students' usual reading strategies (B). It is unlikely that students who receive the reading comprehension strategy (A) first will revert completely to their usual reading habits (B), even when told to do so. Thus, the reading performance under the two conditions is not a fair comparison, because students who receive the AB order are likely to do better under B conditions than students who receive the BA order. That is, intervention A carries over to intervention B, but intervention B does not carry over to intervention A.

Single-case designs. Another category of research designs where participants serve as their own control includes single-case designs (Kazdin, 1982). Although there are many variations of single-case designs, they share the property of repeated measurements of the outcome variable across different conditions, such as before and after the introduction of an intervention or across alternating interventions (an intervention involving changing criteria for performance). Single-case designs offer considerable flexibility because the units of analysis may consist of an individual, several individuals, dyads, small groups, classrooms, or even institutions (Levin, O'Donnell, & Kratochwill, 2003), and when implemented properly, single-case designs can provide strong evidence for causal validity (Kratochwill et al., 2010; Kratochwill & Levin, 2010), although generalizability to the larger population is still limited because the sample size is so small.

In this paper we focus on the multiple-baseline design. It is a flexible single-case design that some argue offers the strongest support for making causal inferences about the effectiveness of interventions (Kratochwill & Levin, 2010). Figure 11 is a schematic representation of one possible multiple-baseline design. There are several prominent features of the multiple-baseline design that are crucial to strengthening internal validity. Its simplicity is that it involves only two phases per case, a baseline phase (A) and an intervention phase (B). A simple AB design with a single case would not control

		Observation														
Case	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	A	A	A	A	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
1	A	A	A	A	A	A	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**
3	A	A	A	A	A	A	A	A	**B**	**B**	**B**	**B**	**B**	**B**	**B**
2	A	A	A	A	A	A	A	A	A	A	**B**	**B**	**B**	**B**	**B**

Note: A = baseline phase; B = intervention phase.

Figure 11. Schematic representation of the multiple-baseline design.

threats to internal validity or support generalization and does not meet evidence standards for making causal claims about an intervention. Notice, however, that the AB phases are replicated across three units or cases (for instance, students). Replication of an intervention effect across cases is a critical requirement for demonstrating experimental control, but it is not sufficient. If the introduction of the intervention occurs at the same time for each case, many threats to internal validity, such as concomitant events, are not controlled. Staggering the introduction of the intervention at different points in time for each individual, however, controls many threats to internal validity. If the outcome measure changes when, and only when, the intervention is introduced, the researcher has evidence of a functional or causal relationship between the intervention and the outcome measure.

Additional control over threats to internal validity can be introduced to the multiple-baseline design via two randomization schemes (Kratochwill & Levin, 2010). One way to introduce randomization into this design is to assign participants at random to the staggered intervention conditions. Additionally, the investigator could assign the timing of the intervention for each case at random. Random assignment, however, leaves open the possibility of having only a single baseline or post-intervention observa-

tion and identical starting points across cases. An alternative approach is the regulated randomization procedure (Koehler & Levin, 1998).

In regulated randomization, the researcher specifies the minimum number of observations required for the baseline and intervention phases of the study by considering the total number of observations available. In Figure 11, a decision was made to have a minimum of three baseline observations with a minimum of five post-intervention observations. Given the 15 available measurement intervals, the intervention could be introduced prior to the fourth observation but no later than after the tenth observation. Accordingly, there are seven possible intervention points that could be selected at random. To maintain clear temporal separation of the introduction of the intervention across cases, regulated randomization allows the researcher to specify a range of starting points for the intervention such that there is no overlap in starting points across cases (Koehler & Levin, 1998). With 15 observation periods available, the researcher could randomly choose to start the intervention prior to the fourth or fifth observation for one case. For the next case, the researcher would use random selection to start the intervention prior to the sixth or seventh observation, and so on for the remaining

two cases. The outcome of the random assignment of starting points for the intervention and random assignment of students to starting points is depicted in Figure 11. If there are more than four cases, students are randomly assigned to one of the four starting points.

In sum, the replication of occasions to observe intervention effects, the temporal separation of starting points of the intervention across replications, the use of random assignment to structure starting points, and the random assignment of cases to starting points strengthen the internal validity of the multiple-baseline design and any resulting claims about a causal relationship between the intervention and the outcome measure of interest.

To meet evidence standards, a multiple-baseline design (and other single-case designs) must meet four criteria (Kratochwill et al., 2010). First, the researcher must document control over how and when the intervention is implemented. Second, it is necessary to establish a high standard of interobserver agreement for the measurement of the outcome measure (greater than 80%). Interobserver agreement must be calculated for each case on 20% of the observations in each phase (baseline and intervention). Third, a multiple-baseline design must include at least three replications of the intervention at three different points in time. Fourth, each phase must include at least five data points. To meet evidence standards with reservations, each phase must include at least three data points. The design in Figure 11 meets all of the criteria except for the last one, because one case has fewer than five observations in one baseline phase. Accordingly, this multiple-baseline design meets evidence standards with reservations. In-creasing the number of observation periods to 18 would enable the researcher to meet the criterion of at least five observations per phase.

To meet these criteria, a study must provide at least three demonstrations of experimental control over the outcome variable (an effect) to claim "strong evidence" of a causal relationship. Otherwise, the study provides "no evidence" to support the claim of a causal relationship. If a study provides three demonstrations of an effect and at least one instance of no effect, then there is "moderate evidence" for a causal relationship (Kratochwill et al., 2010). Additionally, interpretation of the results of a multiple-baseline design must be considered in the context of whether the interventions are defined in terms of meaningful theoretical constructs (for example, scaffolding, peer-tutoring, comprehension monitoring, etc.) and whether the intervention was implemented as intended.

While there is more widespread agreement around the procedures necessary to control threats to the internal validity of the multiple-baseline design, there is considerable debate about the methods for detecting the presence of intervention effects in single-case designs. Visual analysis (Hersen & Barlow, 1976; Kennedy, 2005; Kratochwill et al., 2010); percentage of nonoverlapping data (Scruggs & Mastropieri, 1998, 2013); regression analysis, including hierarchical growth models (Raudenbush & Bryk, 2002, Singer & Willet, 2003); and randomization tests (Edgington, 1975, 1992; Koehler & Levin, 1998; Levin & Wampold, 1999) are three of the primary methods proposed for analyzing time-series data arising from single-case

designs to detect intervention effects. Although a detailed discussion of these topics is beyond the scope of this article, it is important to note that regardless of one's analytical approach there are no clear guidelines about how to integrate findings from single-case designs with studies using more traditional group comparison designs.

Despite the controversy over estimating effect-size estimates from single-case designs, the advantages of a well-controlled single-case design outweigh those of between-group designs, where threats to internal validity are not controlled. Investigators are encouraged to publish the raw data that will enable reviewers to conduct reanalysis of data in the light of new developments. Additionally, effect size estimates from single-case designs can be considered separately from the effect sizes obtained in traditional between-group designs to identify what works. Wendel, Cawthon, Ge, and Beretvas (2015) provide a useful example of a review of studies on individuals who are deaf or hard-of-hearing. As Kratochwill et al. (2010) suggest, the rank ordering of interventions from within-subjects studies and those from between-subjects studies are likely to be similar despite the present incomparability of their effect-size estimates.

Conclusion

In sum, the research designs we have focused on in this article are by no means novel. However, each design offers useful strategies to support causal claims about the effectiveness of interventions with individuals who have sensory disabilities. Undoubtedly, describing strategies for designing research that support causal claims about the effectiveness of interventions among individuals with sensory disabilities is far easier than conducting the studies in the myriad settings in which researchers and teachers work to support them. Progress in the field, finding what works, depends on the ability of the members of the field to make logically defensible arguments for making causal claims about the effectiveness of interventions and the basis for estimating the magnitude of their effects.

We have written this article because we hope to inspire a new generation of researchers to think more creatively about the use of research designs that can strengthen evidence for drawing causal inferences about interventions that promote learning and development among individuals with sensory disabilities. Our reviews have highlighted several difficulties in conducting research with low-prevalence populations, but hopefully they have also suggested methods that can strengthen research in the future. We offer these recommendations to new and experienced researchers, as well as to teachers conducting action research with their students:

1. Research questions should influence the selection of methodology.
2. Researchers should think creatively about how to design research that supports causal validity—and thereby eliminates rival hypotheses.
3. Questions about causal relationships, however, require qualitative descriptions about the nature and fidelity of implementation of interventions.
4. Descriptive statistics that enable other investigators to replicate analyses and synthesize findings of a study,

particularly in relation to other studies examining the same constructs, must be reported. Although journals may request cuts in the manuscript, we believe that these statistics are foundational for a study and vital to its interpretation and application by others.

5. Maximize power by including covariates that reduce error variance in hypothesis testing (see also Wright, 2010).

6. Calculate effect size estimates and confidence intervals for primary study outcomes (Fitz, Morris, & Richler, 2012; Kim, 2015; Peng, Long, & Abaci, 2012).

Many of these recommendations reflect basic principles of research design. However, in the effort to conduct research in low-prevalence fields hampered by populations with small numbers, extreme heterogeneity, and wide geographical dispersion, recognizing what works—and what does not—is often difficult. We hope these recommendations will focus the field on potential solutions.

References

Abou-Gareeb, I., Lewallen, S., Bassett, K., & Courtright, P. (2001). Gender and blindness: A meta-analysis of population-based prevalence surveys. *Ophthalmic Epidemiology*, *8*, 39–56.

Atkinson, R. C. (1968). Computerized instruction and the learning process. *American Psychologist*, *23*, 225–239.

Bartlett, D. J., MacNab, J., MacArthur, C., Mandwich, A., Magill-Evans, J., Young, N. L., Beal, D., Conti-Becker, A., & Polatajko, H. J. (2006). Advancing rehabilitation research: An interactionist perspective to guide question and design. *Disability and Rehabilitation, 28*, 1169–1176.

Botsford, K. D. (2013). Social skills for youths with visual impairments: A meta-analysis. *Journal of Visual Impairment & Blindness*, *107*, 497–508.

Brantlinger, E. V. (2005). Qualitative studies in special education. *Exceptional Children*, *71*, 195–207.

Cawthon, S., & Leppo, R. (2013). Assessment accommodations on tests of academic achievement for students who are deaf or hard of hearing: A qualitative meta-analysis of the research literature. *American Annals of the Deaf, 158*, 363–376.

Conroy, P. W. (2007). Paraprofessionals and students with visual impairments: Potential pitfalls and solutions. *RE:view, 39*, 43–55.

Cook, T. D. (2002). Randomized experiments in educational policy research: A critical examination of the reasons the educational evaluation community has offered for not doing them. *Educational Evaluation and Policy Analysis*, *24*(3), 175–199.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Cronbach, L. J. (1982). *Designing evaluations of educational and social problems*. San Francisco: Jossey-Bass.

Edgington, E. S. (1975). Randomization tests for one-subject operant experiments. *Journal of Psychology*, *90*, 57–68.

Edgington, E. S. (1992). Nonparametric tests for single-case experiments. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis* (pp. 133–157). Hillsdale, NJ: Erlbaum.

Ferrell, K. A. (2006). Evidence-based practices for students with visual disabilities. *Communication Disorders Quarterly*, *28*(1), 42–48.

Ferrell, K. A., Bruce, S., & Luckner, J. L. (2014). *Evidence-based practices for students with sensory impairments* (Document No. IC-4). Gainesville, FL: Collaboration for Effective Educator, Development, Accountability, and Reform Center, University of Florida. Retrieved from http://

ceedar.education.ufl.edu/tools/innovation-configurations

Ferrell, K. A., Buettel, M., Sebald, A. M., & Pearson, R. (2006). *American Printing House for the Blind mathematics research analysis* [Technical Report]. Greeley, CO: University of Northern Colorado, National Center on Low-Incidence Disabilities. Retrieved from http://www.unco.edu/ncssd/research/math_meta_analysis.shtml

Ferrell, K. A., Dozier, C., & Monson, M. (2011). *Meta-analysis of the educational research in low vision*. Greeley, CO: University of Northern Colorado. Retrieved from http://www.unco.edu/ncssd/research/LowVisionMeta-Analysis.shtml

Ferrell, K. A., Mason, L., Young, J., & Cooney, J. (2006). *Forty years of literacy research in blindness and visual impairment* [technical report]. Greeley, CO: University of Northern Colorado, National Center on Low-Incidence Disabilities. Retrieved from http://www.unco.edu/ncssd/research/literacy_meta_analyses.shtml

Fitz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates, current use, calculations, and interpretation. *Journal of Experimental Psychology, 141*, 2–18,

Forster, E. M., & Holbrook, M. C. (2005). Implications of paraprofessional supports for students with visual impairments. *RE:view, 36*, 155–163.

Gersten, R. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*, 149–164.

Graeme, D., McLinden, M., McCall, S., Pavey, S., Ware, J., & Farrell, A. M. (2011). Access to print literacy for children and young people with visual impairment: Findings from a review of literature. *European Journal of Special Needs Education, 26*, 25–38.

Griffin-Shirley, N., & Matlock, D. (2004). Paraprofessionals speak out: A survey. *RE:view, 36*, 127–136.

Harris, B. (2011). Effects of the proximity of paraeducators on the interactions of braille readers in inclusive settings. *Journal of Visual Impairment & Blindness, 105*, 467–478.

Hersen, M., & Barlow, D. H. (1976). *Single-case experimental designs: Strategies for studying behavior change*. New York: Pergamon Press.

Horner, R. M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–179.

Individuals with Disabilities Education Act, 20 U. S. C. § 1400 (2004).

Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. (2008). *Technical methods report: Statistical power for regression discontinuity designs for education evaluation*. Washington, DC: U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/pubs/20084026/index.asp

Joint Committee on Infant Hearing. (2007). Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. *Pediatrics, 120*, 898–921.

Kazdin, A. E. (1982). *Single-case research designs: Methods for clinical and applied settings*. New York: Oxford University Press.

Kelly, S. M., & Smith, D. W. (2011). The impact of assistive technology on the educational performance of students with visual impairments: A synthesis of the research. *Journal of Visual Impairment & Blindness, 105*, 73–83.

Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Allyn and Bacon.

Keppel, G., & Zedeck, S. (1989). *Data analysis for research designs*. New York: W. H. Freeman.

Kim, D. S. (2015). Power, effect size, and practical significance: How the reporting in *Journal of Visual Impairment & Blindness* articles has changed in the past 20 years. *Journal of Visual Impairment & Blindness, 109*, 214–218.

Koenig, A. J., & Holbrook, M. C. (2000). Professional practice. In M. C. Holbrook & A. J. Koenig (Eds.), *Foundations of edu-*

cation: Vol. 1. History and theory of teaching children and youths with visual impairments* (2nd ed., pp. 260–276). New York: AFB Press.

Koehler, M. J., & Levin, J. R. (1998). Regulated randomization: A potentially sharper analytical tool for the multiple-baseline design. *Psychological Methods*, *3*(2), 206–217.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*, *15*(2), 124–144.

Levin, J. R., & Wampold, B. E. (1999). Generalized single-case randomization tests: Flexible analyses for a variety of situations. *School Psychology Quarterly*, *14*, 59–93.

Levin, J. R., O'Donnell, A. M., & Kratochwill, T. R. (2003). Educational/psychological intervention research. In I. B. Weiner (Series Ed.), W. M. Reynolds, & G. E. Miller (Vol. Eds.), *Handbook of psychology: Vol. 7. Educational psychology* (pp. 557–581). New York: Wiley.

Lewis, S., & McKenzie, A. R. (2010). The competencies, roles, supervision, and training needs of paraeducators working with students with visual impairments in local and residential schools. *Journal of Visual Impairment & Blindness, 104*, 464–477.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Luckner, J. L. (2006). Evidence-based practices with students who are deaf. *Communication Disorders Quarterly*, *28*, 49–52.

Luckner, J. L., & Cooke, C. (2010). A summary of the vocabulary research with students who are deaf or hard of hearing. *American Annals of the Deaf*, *155*, 38–67.

Luckner, J. L., & Handley, C. M. (2008). A summary of the reading comprehension research undertaken with students who are deaf or hard of hearing. *American Annals of the Deaf*, *153*, 6–36.

Luckner, J. L., Sebald, A. M., Cooney, J., Young, J., & Muir, S. G. (2005, 2006). An examination of the evidence-based literacy research in deaf education. *American Annals of the Deaf, 150*, 443–456.

Luckner, J. L., & Urbach, J. (2012). Reading fluency and students who are deaf or hard of hearing: Synthesis of the research. *Communication Disorders Quarterly*, *33*, 230–241.

Marks, S. U., Schrader, C., & Levine, M. (1999). Paraeducator experiences in inclusive settings. Helping, hovering, or holding their own? *Exceptional Children, 65*, 315–328.

McKenzie, A. R., & Lewis, S. (2008). The role and training of paraprofessionals who work with students who are visually impaired. *Journal of Visual Impairment & Blindness, 102*, 459–471.

Meinzen-Derr, J., Wiley, S., & Choo, D. I. (2011). Impact of early intervention on expressive and receptive language development among young children with permanent hearing loss. *American Annals of the Deaf, 155*(5), 580–591.

Moeller, M. P. (2000). Early intervention and language development in children who are deaf and hard of hearing. *Pediatrics, 106*(3), E43.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107–110, § 115, Stat. 1425 (2002).

Odom, S. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, *71*, 137–148.

Parker, A. T., Davidson, R., & Banda, D. R. (2007). Emerging evidence from single-subject research in the field of deaf-blindness. *Journal of Visual Impairment & Blindness, 101*, 690–700.

Parker, A. T., Grimmett, E. S., & Summers, S. (2008). Evidence-based communication practices for children with visual impairments and additional disabilities: An examination of single-subject design studies.

Journal of Visual Impairment & Blindness, 102, 540–552.

Parker, A. T., & Ivy, S. (2014). Communication development of children with visual impairment and deafblindness: Synthesis of intervention research. In D. D. Hatton (Ed.), *International Review of Research in Developmental Disabilities, 46*, 101–144.

Parker, A. T., & Pogrund, R. L. (2009). A review of research on the literacy of students with visual impairments and additional disabilities. *Journal of Visual Impairment & Blindness, 103*, 635–648.

Peng, C.-Y. J., Long, H., & Abaci, S. (2012). Power analysis software for educational researchers. *The Journal of Experimental Education, 80*, 113–136.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd Ed). Thousand Oaks, CA: Sage Publications.

Rivlin, A. M., & Timpane, P. M. (1975). *Planned variation in education: Should we give up or try harder?* Washington, DC: Brookings Institution.

Russotti, J., & Shaw, R. (2001). Inservice training for teaching assistants and others who work with students with visual impairments. *Journal of Visual Impairment & Blindness, 95*, 483–487.

Scruggs, T. E., & Mastropieri, M. A. (1998). Summarizing single-subject research: Issues and applications. *Behavior Modification, 22*, 221–242.

Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education, 34*(1), 9–19.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton-Mifflin.

Shavelson, R. J., Towne, L., & the Committee on Scientific Principles for Education Research. (Eds.). (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.

Snyder, T. D., & Dillow, S. A. (2015). *Digest of Education Statistics, 2013* [Table 204.30]. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2015011

Sullivan, G. M. (2011). Getting off the "gold standard": Randomized controlled trials and education research. *Journal of Graduate Medical Education, 3*, 285–289. doi: 10.4300/JGME-D-11-00147.1

Thompson, B. W. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children, 71*, 181–194.

Thompson, B., Diamond, K. E., McWilliam, R., Snyder, P., & Snyder, S. W. (2005). Evaluating the quality of correlational research for evidence-based practice. *Exceptional Children, 71*(2), 195–207.

Valentine, J. C., & Cooper, H. (2004). *What Works Clearinghouse study design and implementation assessment device* (Version 1.1). Washington, DC: Institute of Education Sciences, U.S. Department of Education.

Valentine, J. C., & Cooper, H. (2008). A systematic and transparent approach for assessing the methodological quality of intervention effectiveness research: The Study Design and Implementation Assessment Device (Study DIAD). *Psychological Methods, 13*, 130–149. Retrieved from http://dx.doi.org/10.1037/1082-989X.13.2.130

Vlastarakos, P. V., Proikas, K., Papacharalampous, G., Exadaktylou, I., Mochloulis, G., & Nikolopoulos, T. P. (2010). Cochlear implantation under the first year of age—The outcomes. A critical systematic review and meta-analysis. *International Journal of Pediatric Otorhinolaryngology, 74*, 119–126.

Vohr, B., Jodoin-Krauzyk, J., Tucker, R., Johnson, M. J., Topol, D., & Ahlgren, M.

(2008). Results of newborn screening for hearing loss: Effects on the family in the first 2 years of life. *Archives of Pediatrics & Adolescent Medicine, 162*(3), 205–211.

Wang, Y., & Williams, C. (2014). Are we hammering square pegs into round holes? An investigation of the meta-analyses of reading research with students who are deaf or hard of hearing and students who are hearing. *American Annals of the Deaf, 159,* 323–345.

Wauters, L. N. (2001). Sign facilitation in word recognition. *Journal of Special Education, 35*(1), 31–40.

Wendel, E., Cawthon, S. W., Ge, J. J., & Beretvas, S. N. (2015). Alignment of single-case design (SCD) research with individuals who are deaf or hard of hearing with the What Works Clearinghouse Standards for SCD Research. *Journal of Deaf Studies and Deaf Education Advance Access, February,* 1–12. doi:10.1093/deafed/enu049

What Works Clearinghouse. (2014). *Procedures and standards handbook version 3.0.* Washington, DC: Institute of Education Sciences.

Wright, T. (2010). Looking for power: The difficulties and possibilities of finding participants for braille research. *Journal of Visual Impairment & Blindness, 104,* 775–780.

Wright, T., Harris, B., & Sticken, E. (2010). A best-evidence synthesis of research on orientation and mobility involving tactile maps and models. *Journal of Visual Impairment & Blindness, 104,* 95–106.

Yoshinaga-Itano, C. (2003a). Early intervention after universal neonatal hearing screening: Impact on outcomes. *Mental Retardation and Developmental Disabilities Research Reviews, 9,* 252–266.

Yoshinaga-Itano, C. (2003b). From screening to early identification and intervention: Discovering predictors to successful outcomes for children with significant hearing loss. *Journal of Deaf Studies and Deaf Education, 8,* 11–30.

Yoshinaga-Itano, C., & Gravel, J. S. (2001). The evidence for universal newborn hearing screening. *American Journal of Audiology, 10,* 62–64.

Yoshinaga-Itano, C., Sedey, A. L., Coulter, D. K., & Mehl, A. L. (1998). The language of early- and later-identified children with hearing loss. *Pediatrics, 102,* 1161–1171.

John B. Cooney, Ph.D., associate vice president for administration, Office of Academic Affairs, University of Colorado, Campus Box 35 UCA, 1800 Grant Street, Suite 800, Denver, CO 80203; e-mail: <john.cooney@cu.edu>. John Young III, Ph.D., vice president of development and data science, Applied Analytics Group, DST Systems, 333 West 11th Street, Kansas City, MO 64105; e-mail: <outcome32@yahoo.com>. John L. Luckner, Ed.D., professor of special education and coordinator of the Deaf Education Program, College of Education and Behavioral Sciences, McKee 29, Campus Box 141, University of Northern Colorado, Greeley, CO 80639; e-mail: <john.luckner@unco.edu>. Kay Alicyn Ferrell, Ph.D., professor emerita of special education, School of Special Education, College of Education and Behavioral Sciences, University of Northern Colorado; e-mail: <kay.ferrell@bears.unco.edu>. Please address all correspondence to Dr. Cooney.