Single-Case Experimental Research: A Methodology for Establishing Evidence-Based Practice in Special Education

Faisl Alqraini,
Department of Special Education,
Prince Sattam bin Abdulaziz University,
Saudi Arabia

Abstract
In the field of special education there is a dearth of group experimental studies that establish evidence-based practice. The effort to establish evidence-based practice has been associated with emphasizing experiments by using randomized controlled trial with large numbers of participants who are randomly assigned to a treatment. However, single-case design (SCD) research can play a vital role in filling the gaps and determining educational interventions that establish evidence-based practices in the special education. The goal of this present paper is to provide an overview of SCD methods and how these methods can establish evidence-based practices in special education. The author shares the critical features of SCD, including the way steady state strategy and baseline logic works in common designs. Internal and external validity are also addressed. Finally, this paper communicates why a visual analysis of data is considered conservative in SCD rather than tests of statistical significance.

Keywords: *Special Education; Single-Case Design; Evidence-Based Practice*

Introduction

Applying high quality research is essential to establish an evidence-based practice in special education. Most evidence-based practices in special education emphasize establishing an intervention, instructional strategy, or teaching program based on a scientific approach

(Mesibov & Shea, 2011). Organizations in general education and special education such as the What Works Clearinghouse, Council for Exceptional Children, and Division 16 of the American Psychological Association have created guidelines to determine quality and quantity of research that leads to evidence-based practice (Plavnick & Ferreri, 2013). The reason behind this educational policy is that practices with prior evidence of effectiveness under research conditions are more likely to produce positive educational outcomes (Odom et al. 2005).

The scientific community reviews randomized controlled trials as the "*gold standard*" to establish evidence-based practice in intervention research (Kazdin, 2011; Plavnick & Ferreri, 2013). In special education, it is a quite difficult to apply randomized controlled trials because students' characteristics vary across multiple factors. Cannon, Guardino, Antia, and Luckner (2016) assert that deaf and hard of hearing students vary across multiple factors such as, a) hearing level, b) age at identification, c) age at amplification, d) language exposure, e) early intervention, f) additional eligibility for specialized services, and g) parental hearing status. SCD could benefit special education researchers in establishing evidence-based practice. The goal of this paper is to provide an explanation of the ways in which SCD can be used to document evidence-based practices in special education. First, it presents an overview of the critical features of SCD, how internal and external validity are addressed in SCD, along with an explanation regarding the way in which a steady-state strategy and baseline logic work in common design such as reversal and multiple baseline experimental designs. Second, the paper describes the way to determine a functional relationship (to document causal) in such designs. It also discusses why visual analysis of data rather than tests of statistical significance is considered a conservative approach in SCD.

Even though a number of textbooks discuss single-case design deeply and thoroughly, these textbooks discuss single-case design in a broader context. Adding to this, many textbooks do not address applying single-case design in special education field. That is, they do not use relevant examples in how these designs can be used in special education settings. Examples that illustrate single-case design in this paper were selected from published literature in special education.

Features of SCD Methodology

SCDs are true experiments, and thus they can determine the causal relationship between independent and dependent variables (Horner et al., 2005; Kazdin, 2011). The distinctive characteristic of SCDs is their ability to evaluate interventions with just one or a few participants. Results from group data definitely are essential; however, researchers seek to understand, with greater specificity, the effectiveness of interventions on participants as individuals, which group studies typically do not address. Therefore, SCDs can be used to improve an intervention in case it does not work or fails to achieve the goal (Kazdin, 2011). SCD has several critical features that make it distinctive from other designs. Some key features are: 1) the individual participant as the unit of analysis, 2) the operational definition of study characteristics, 3) the use of baseline/intervention conditions, 4) experimental control, 5) the repeated measurement of target behaviors, 6) the repeated and systematic introduction of interventions, and 7) visual analysis (Cakiroglu, 2012; Horner et al., 2005). These features contribute to establishing evidence-based practices in special education. Each feature is described in the following sections. The use of baseline/intervention conditions, and experimental control were discussed deeply in steady state strategy and baseline logic section.

1.) The individual participant as the unit of analysis

SCD can only use one participant; however, more than one participant (e.g., 3 to 8) is desirable in SCD study. Each individual serves as his/her own control. Performance prior to intervention is compared to performance during and/or after intervention (Horner et al., 2005). Therefore, researchers use individual data analysis to show the effectiveness of an intervention.

2.) Operational definition of study characteristics

Defining aspects such as the target behavior(s), participant demographics, and research setting, are pivotal in SCD studies because these explicit definitions allow other researchers to more accurately replicate an experiment (Cakiroglu, 2012; Kratochwill et al., 2010). For example, in deaf education, specific characteristics that should be defined and documented in detail, include: a) degree of hearing loss, b) age at onset of hearing loss, c) gender, d) ethnicity, e) presence of additional identified disabilities, f) mode of communication (speech, sign, etc.), g) , and h) hearing and socioeconomic status (SES) of parents/caregivers (Paul, Wang, & Williams, 2013). If the researcher does not clearly and objectively define the target behavior or skill, replication is less likely (Cooper, Heron, & Heward, 2007). The definition of dependent variables should be objective, observable and measurable in order to help reduce any likely disagreement between observers collecting data (Sealander, 2004). Without directly well-written definitions, researchers cannot be able to clearly and truthfully measure the target behavior. Table 1 illustrates some operational definitions for dependent variables.

Table 1. Examples of operational definitions for dependent variables

Behavior	**Definition**
Word Recognition (Dolch words and bridge phrases).	"The teacher signed each word or phrase to the student and asked him or her to label the word or phrase by fingerspelling the word, voicing the word, or pointing to the word that corresponded to the sign produced by the researcher" (Dimling, 2010, p 430).
Reading Comprehension	“The operational definition for reading comprehension [is] the number of details students retold in one-minute, following oral reading of instructional-level, content-area text” (Benedict, Rivera &Antia, 2015 p. 3).

In addition to this, investigations must indicate for whom the practice is effective and in what context” (Odom et al. 2005).

3.) Repeated and systematic measurement of target behaviors

Repeated measurement of target behaviors illustrates the performance of a participant in both the baseline and the intervention conditions (Kazdin, 2011). The baseline is the first phase in SCD and is used to establish initial patterns of behavior. The baseline will be utilized later in order to compare the performance of a participant after an intervention is introduced (Kennedy, 2005). The intervention phase is implemented after the baseline phase when the baseline has been established as relatively stable. Systematic repeated measurements reveal that the data was continuously recorded on a target behavior, thus providing a true representation of the participant’s performance during each session. In group design, the only way to observe participant’s performance is by using pretests and post-tests and comparing the results to control group. Whereas in single-case design, a participant’s performance can be observed during each session, which means controlling to any threats to internal (Neuman & McCormick, 1995; Tankersley, Harjusola-Webb & Landrum, 2008). In addition, the use of repetition illustrates the casual relationship between the dependent and the independent variable. (how this occurred discussed below)

4.) Visual Analysis

Visual analysis is the main approach to analyzing data in single-case design (Kahng et al., 2010). The visual analysis shows the basis for comparing pre-intervention and intervention phases from baseline phase to intervention phase (Kratochwill & Levin, 2010). Six critical elements are used to assess progress within- and between-phase data patterns: a) level (mean), b) trend (slope), c) variability (standard deviation), d) immediacy of the effect, e) overlap, and f) consistency of data patterns across similar phases (Kratochwill et al., 2010). Visual analysis of these elements is used to determine whether a causal relationship exists between an independent variable and a dependent variable.

Internal and External Validity

Researchers implementing intervention studies are concerned with internal and external validity (Cannon, Guardino, Antia, & Luckner, 2016). *Internal validity* means that the intervention changes the dependent variable and extraneous factors are not influenced by the outcome of the intervention. When the results are attributed with little or no vagueness to the impact of the intervention, the experiment is deemed to be internally valid (Kazdin, 2011). Threats to internal validity can be enhanced through replication (Kratochwill et al., 2010). *External validity* refers to the extent the results of the intervention can be generalized to other people and contexts beyond those who participated in the study (Horner et al., 2005). External validity can be improved through replication.

Undoubtedly, one or few participants is not representative of all individuals coming from the same population. In this case, researchers can maintain external validity by replicating the intervention. Generalization can be confirmed by direct and systematic replications. In direct replication, the researcher repeats the same study (study characteristics) with participants who have similar characteristics, whereas systematic replication is repeating the study with participants who have different characteristics than those who participated in the original study. In the field of special education, the researcher can conduct a study in an elementary school and then replicate it with students in high school (Cakiroglu, 2012). Major threats to internal validity can be controlled over steady state strategy and baseline logic.

Steady State Strategy and Baseline Logic

Steady state strategy is defined as "an approach to making experimental comparisons that involves measuring responding for each participant repeatedly under each condition in an effort to assess and manage extraneous influences and thereby obtain a stable pattern of responding that represents the full effects of each condition" (Johnston & Pennypacker, 2009, p. 195). Steady state strategy shows the basis for baseline logic as can be seen in Figure 2.

Baseline Logic

Baseline logic encompasses three factors: prediction, verification, and replication Cooper et al.,
2007). However, some scholars such as Riley-Tillman and Burns (2009) add *affirmation of the consequent* as a separate phase. Affirmation of the consequent is classified as a form of replication in traditional experiments (Alberto & Troutman, 2013). These guide the baseline logic and present them visually that can be seen in figure 5.

Baseline Data

The first phase in SCD encompasses gathering and recording of baseline data of the level of the dependent variable in the natural environment before applying the intervention for several days (Alberto & Troutman, 2013). Therefore, the baseline data or phase serves two critical functions. First, baseline data serves as a *descriptive function* whereas the data depicts the existing level of performance. Second, baseline data serves as a *predictive function* whereas the data serves as a predication of the level of performance before an intervention is implemented (Kazdin, 2011).

Predication.

The baseline phase needs to be continuous for many sessions prior to the implementation of the intervention in order to judge the effectiveness of the intervention. Also, it is critical that the data be stable. The stability can be assessed by two factors: variability and trends. Simply put, variability of data means oscillating student's performance through the baseline phase when the variability in data becomes highly oscillatory, which makes it more difficult to draw conclusions regarding the effects of the intervention. In this case, when the baseline is unstable, a researcher needs to look at the definition of a target behavior because the operational definition of target behavior may not adequately and accurately be described, table 1 showed a perfect example of a good operational definition of target behavior. Trend in data points involves three successive data in the same direction, which means the baseline may present no trend, increase trend, or decrease trend, as seen in Figure 1. An ascending baseline indicates an increasing trend; therefore, in this case the researcher can apply the intervention that aims to decrease the behavior, and vice versa with a descending baseline (Alberto & Troutman, 2013).

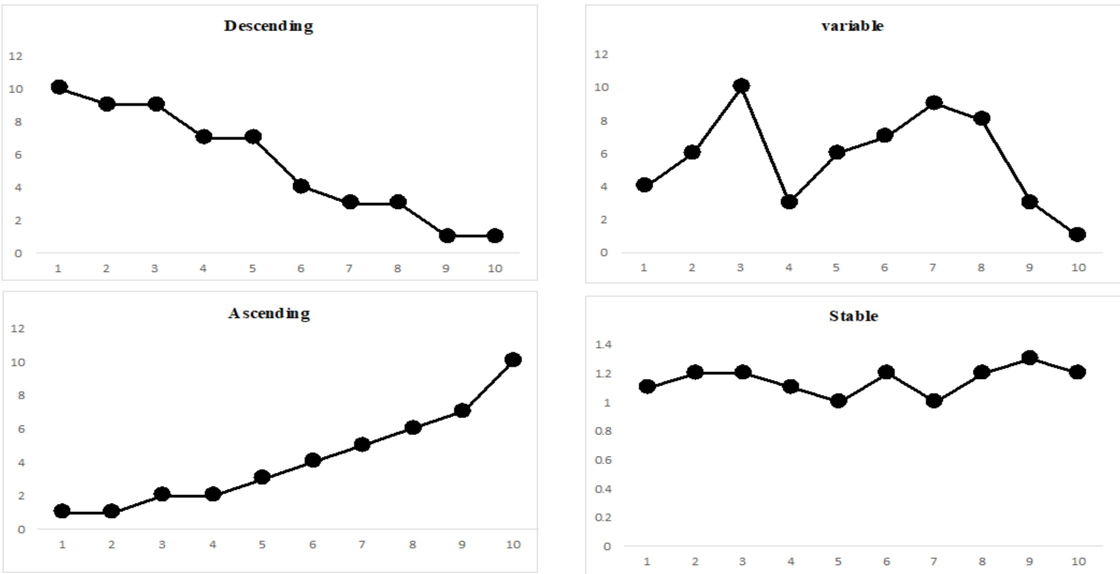

Figure 1. An example of descending, ascending, variability, and stability of data.

Affirmation of Consequent.

Essentially, the independent variable has to be implemented when the baseline is stable in order to see whether the intervention results in a change in the behavior or not. This is referred to as *affirmation of the consequent.* Figure 2 shows a successful *affirmation of the consequent.* Steady state responding over the baseline allows the forecast if the treatment was not implemented (or not affective), then the data will still be the same as the data in the shaded area. Thus, the intervention was implemented and we can clearly see how the target behavior was changed repeatedly. This is detecting the correlation between the independent and dependent variables.

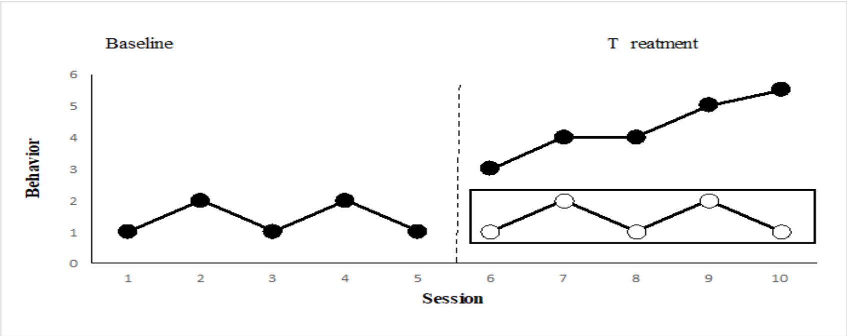

Figure 2. Baseline logic: Affirmation of the consequent.

Verification.

The verification phase is conducted in order to confirm the initial hypothesis of the previous predication baseline. For example, removing the intervention so that the target behavior returns back to baseline level is an example of verification (Riley-Tillman & Burns, 2009). In other words, when the researcher wants to double check that the intervention was influential to the target behavior after the predication baseline was stable, then the researcher removes the intervention. If the intervention results in a decrease of target behavior, then the predication is verified. Figure 3 illustrates the procedure.

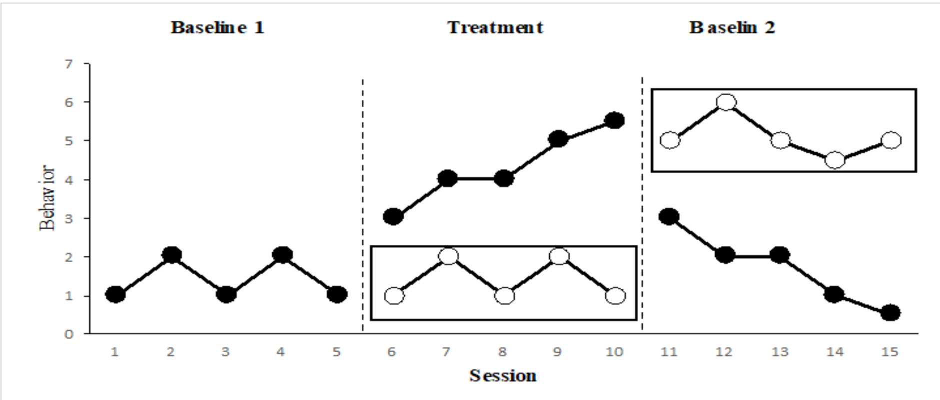

Figure 3. Baseline logic: Verification.

Replication.

"Replication is the essence of believability "(Baer, Wolf, & Risley, 1968, p. 95).
The goal of the replication phase is to confirm the initial intervention effects observed in the affirmation of the consequent phase (Riley-Tillman & Burns, 2009). After steady state responding is achieved over baseline 2, the intervention is reintroduced over intervention 2 phase in order to determine the reliability of a functional relationship between the independent and dependent variables. Also, it offers controlling of all extraneous factors that might effect treatment.

Types of SCD

Reversal, alternating, multiple baseline, and changing criterion designs are all types of SCD. However, this paper highlights only two designs (reversal and multiple baseline designs) because they are most commonly used in special education published research. Hammond and Gast (2010) review eight special education journals between 1983 and 2007 to identify which methods are commonly used. 1,936 articles were reviewed, and the researchers concluded that reversal designs and multiple baseline designs were commonly used compared to other designs.

Reversal Design.

Reversal or withdrawing design is widely used by special education researchers. Therefore, it entails repeated measures of behavior in three phases sequentially, at least: a) an initial baseline phase, that is, the independent variable (intervention or practice) is not presented, b) an intervention phase, that is, the intervention is presented, and c) a return to baseline condition, that is, the intervention or practice is removed in order to see to what extent the intervention is effective (Cooper et al., 2007). Consequently, reversal design refers to the withdrawal of an independent variable during one or more phases of a study in order to explain a practical relationship between the independent variable and the target behavior (Richards et al., 1999). As seen, in figure 4, baseline (A) denotes the data are collected until the steady state responding has been achieved. The following step is the treatment phase (B) wherein the intervention is applied. Then, return to (A) phase in which the intervention is removal in order to demonstrate the functional relationship between the intervention and the target skill (Cakiroglu, 2012). This is called **A-B-A design,** which is the basis of reversal design. In addition, there are other types of design, such as A-B-A-B; A-B-A-B-A-B; B-A-B and so on.

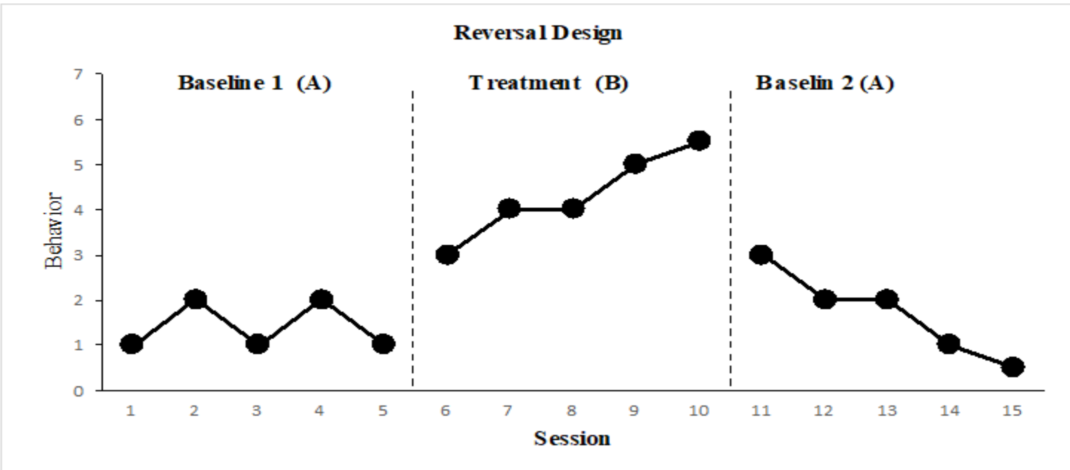

Figure 4. A-B-A design.

The ideal reversal design is an ABAB paradigm due to reintroducing B phase, which means there is a replication of the intervention effects to demonstrate the change in the target skill. This design is straightforward and powerful in demonstrating effective relationships between an independent and dependent variable (Cooper et al., 2007). The purpose, and thus, operation of ABAB design is that: the baseline (A) data are gathered on a target skill for few days. Then, the researcher initiates the treatment (B) for a certain time and gathers data on the same target skill. After that, the researcher returns to (A) phase in which she/he removes the treatment to identify whether or not the target skill comes back to the initial baseline data level. Eventually, the treatment is reintroduced to confirm the alteration in the target skill that was forecast (B), see figure 5 (Risley. 2005). This paradigm is appropriate in terms of increasing or decreasing a target skill that is flexible or malleable such as problem behavior. For example, Belfiore, Basile, and Lee (2008) implemented an intervention by using A-B-A-B reversal design with a 7-year-old, first grade Caucasian boy with Down syndrome, who was also diagnosed with attention deficit hyperactivity disorder (ADHD). The research aimed to determine the effect of the high-probability command sequence (HPCS) on student compliance to low probability requests. The student was frequently non-compliant to classroom requests by saying “no,” or ignoring the requests. He was enrolled in a life skills program located in a general education elementary school with other students with developmental disabilities. The dependent variable was percentage compliance to a low probability command (e.g., "*Go to your desk,*" "*Sit down.*"). Compliance to the HPCS (e.g., "*Touch head,*" "*Give me high five,*" "*Clap your hands*") was also observed to assure the command requests stayed at a high level of compliance throughout the study.

In first baseline phase (A), mean compliance to low probability commands was 13%, and ranged from 0–30% compliance. Through the first intervention phase (B), compliance immediately increased, and stayed high at a mean of 78%, and ranged from 70–90%. When the researchers removed the intervention, the data immediately decreased and stabilized at levels similar to the first baseline. Through the return to baseline (A), mean compliance to low probability commands returned to a low level of 17% compliance, ranging from 10% to 30% compliance. When the intervention phase was re-introduced (B), mean compliance increased to 85%, ranging from 80% to 90% compliance. Through the follow-up phase, 7 days after the intervention was completed, compliance to low-probability commands stayed high at 90% for both sessions.

The role of steady state strategy (predication, verification, and replication) was illustrated in figure 5. As illustrated here, after a stable pattern of responding is achieved through baseline 1, the treatment is introduced. Hypothetically, if the treatment is not introduced or simply not effective, it will still be the same, as suggested in the data in the rectangle shape in treatment 1; this shows predication part. Next, when steady state responding is obtained in Treatment 1, the intervention is removed, and the baseline reverts. When baseline 2 is the same or roughly the same as baseline 1, verification of predication is obtained. Also, when an intervention is reintroduced and given the same as treatment 1 or approximately the same, replication is achieved and we can assert that the intervention works (Cooper et al., 2007).

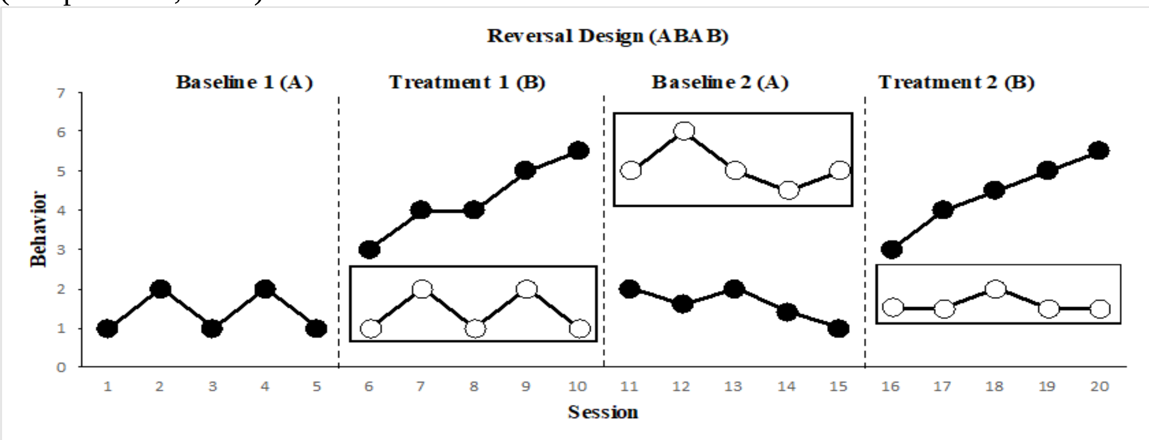

Figure 5. A-B-A-B design.

Types of the A-B-A-B design

Repeated Reversals or Multiple Reversals.

Simply put, this model is an extension of A-B-A-B design, in which the independent variable is removed and reintroduced a second time like so A-B-A-B-A-B (see figure, 6). A researcher who uses multiple reversals shows convincing evidence of the functional relationship between an independent and dependent variable (Cooper et al., 2007).

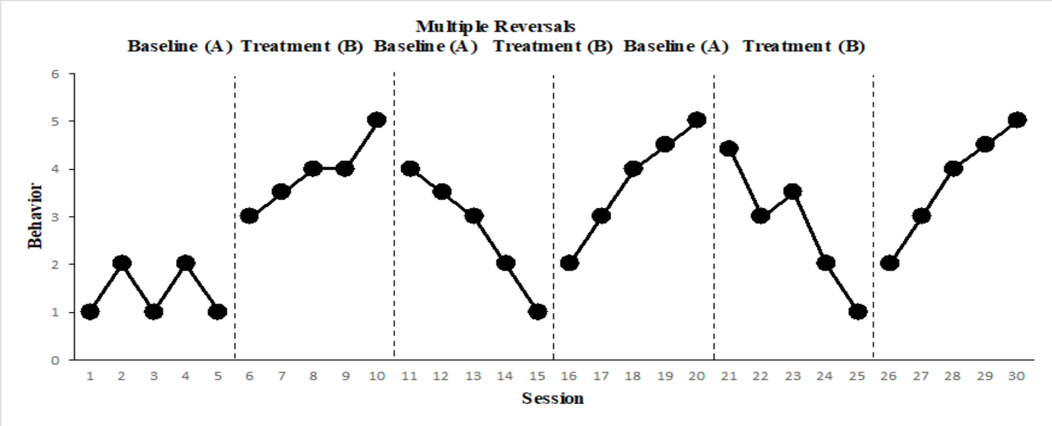

Figure 6. Multiple reversals.

B-A-B Design

B-A-B design reverses A-B-A design, that is, the independent variable is implemented in the first phase (B), then removed in phase (A), and then reinstated again. B-A-B design is a weak design due to not being able to assess the impact of treatment on the pre-intervention level of responding. The non-treatment in phase (A) cannot confirm a prediction of the previous nonexistent baseline. Therefore, B-A-B design presents no data to show "whether the measures of behavior taken during the A condition represent preintervention performance, sequence effects cannot be ruled out: The level of behavior observed during the A condition may have been influenced by the fact that the treatment condition preceded it" (Cooper et al., 2007, p. 180).

Multiple Treatment Reversal Designs.

Multiple treatment reversal designs are used to compare the impact of two or more interventions on the baseline condition or to compare interventions with each other (e.g., A-B-A-C-A-C, A-B-A-C-B-C-B, A-B-C-B-C-B-C, and A-B-A-C-A-D-A-C-A-D) (Hammond & Gast, 2010). The letters *C*, *D*, and so on denote additional conditions (Cooper et al., 2007). Therefore, a researcher who uses reversal design to compare two or more treatments are more likely to encounter two extraneous variables by sequence effects that must be taken into account in order to control them as much as possible.

Implication into Reversal Designs

There are some behaviors that cannot be implemented with reversal design such as learning to read. As children learn how to read or decode certain words, it is impossible for them to get back to the previous phase, that is, prior to the acquisition of skills. Consequently, a researcher cannot show a functional relationship between independent and dependent variables because the target skill is not revertible (Cakiroglu, 2012). However, reversal design can be used with problematic behaviors as was demonstrated in the Belfiore, Basile, and Lee (2008) study.

Multiple Baseline Designs

The multiple baseline design is the most common used experimental design with regard to evaluating intervention effectiveness. It is a powerful strategy that allows researchers to examine the impact of an independent variable through multiple settings, behaviors, and participants without resorting to removing the intervention so that it verifies the development in behavior (Cooper et al., 2007). As discussed previously in reversal design, the nature of reversal design requires the removal of the treatment in order to verify the prediction that was established in the baseline condition. However, the equation is different in the multiple baseline experimental design.

There are three basic types of multiple baseline designs:

- Multiple baseline across behaviors, that is, it encompasses two or more different treatments of the same participant or participants.
- Multiple baseline across settings, that is, there is the same treatment with the same participant in different settings, situations, or time periods.
- Multiple baseline across subject, that is, it encompasses the same treatment with two or more participants (Cooper et al., 2007).

In a multiple baseline design, researchers combine a baseline condition with a treatment condition across participants, behaviors, or settings. A stable baseline is obtained for a participant, behavior, or setting, then the treatment is implemented for the first participant, behavior, or setting. The data are gathered for another participant, behavior, or setting. When the alteration occurs for the baseline that receives the treatment, the procedure is repeated with the second behavior, participant, or setting (Tankersley et al., 2008).

To understand the procedure more clearly, figure 7 presents an illustrative picture of multiple baseline designs. As can be seen, the data shows predicted measures if the baseline condition is not changed (rectangle area in behavior 1, 2, and 3). Baseline data points in both behavior 2 and 3 with bracket demonstrates the prediction of behavior 1. Whereas behavior 3 with bracket A demonstrates the prediction of behavior 2, data obtained through the intervention condition for both behavior 1 and 2 as can be seen in both B brackets, provide replications of the effectiveness of the intervention (Cooper et al., 2007).

For example, Dimling, (2010) applied a multiple baseline design to determine the effects of the vocabulary intervention on word recognition, production, and comprehension with six 2nd graders who were deaf. Hearing losses were all of a bilateral sensorineural nature and ranged from moderate to profound. Two of the 6 students had additional disabilities. Two types of vocabulary words were used for the vocabulary intervention: adapted Dolch words (commonly found in most basal readers) and bridge phrases (e.g., fall down and clean up). This intervention last 6 weeks and all of the students were taught 12 words each week. The results indicated the baseline means for mastered Dolch word recognition, production, and comprehension ranged from 0.00 to 0.33 out of 3.00 words mastered. The baseline means for mastered bridge phrase recognition, production, and comprehension ranged from 0.00 to 0.20 out of 3.00 phrases. All six students had difficulty mastering both types of vocabulary (i.e., Dolch words and bridge phrases) across all three variables (i.e., recognition, production, and comprehension).

In the baseline phase, two students remained in baseline for 3 sessions, two students remained in baseline for 5 sessions, and two students remained in baseline for 7 sessions. The first group of students, who remained in baseline for only 3 sessions, received an intervention in session 4 while the other groups remained in baseline without an intervention in order to see if the intervention was actually effective with the first group or not. When the first group showed a significant increase in dependent variables, the second group received the intervention in order to replicate the effectiveness of the intervention that occurred with first group. The second group was immediately effected when the intervention was introduced, while the third group remained in baseline phase. Finally, the third group was introduced to the intervention after the second group showed improvement in all three variables. This kind of staggered implementation demonstrated how the intervention was effective, as the data revealed the same positive results with all three groups (six participants). The results of the intervention showed that students mastered 78%-100% of the Dolch words and 5%-97% of the bridge phrases over the course of the vocabulary intervention.

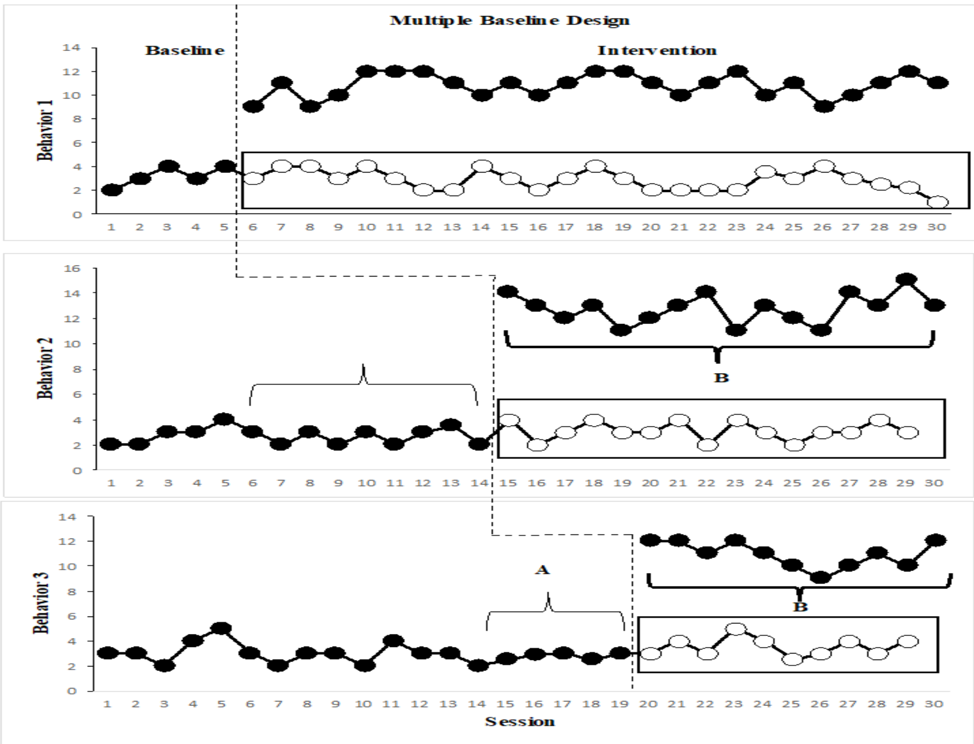

Figure 7. Multiple baseline design.

Design Variation

Multiple Probe Designs.

In the multiple probe design, sporadic measures occur at the initial stage of the experiment and then after each time a subject has mastered one of the behaviors or sequential skills. True baselines are gathered for each behavior prior to instruction (intervention). The multiple probe design is beneficial for assessing the impact of instruction on skill sequences when it is not likely to happen that the subject will master later steps without instruction. In addition, multiple probe design is also beneficial for cases where a lengthy baseline could have negative affects for the subject or the intervention (Cooper et al., 2007). Figure 8 offers an example of how this design can be implemented.

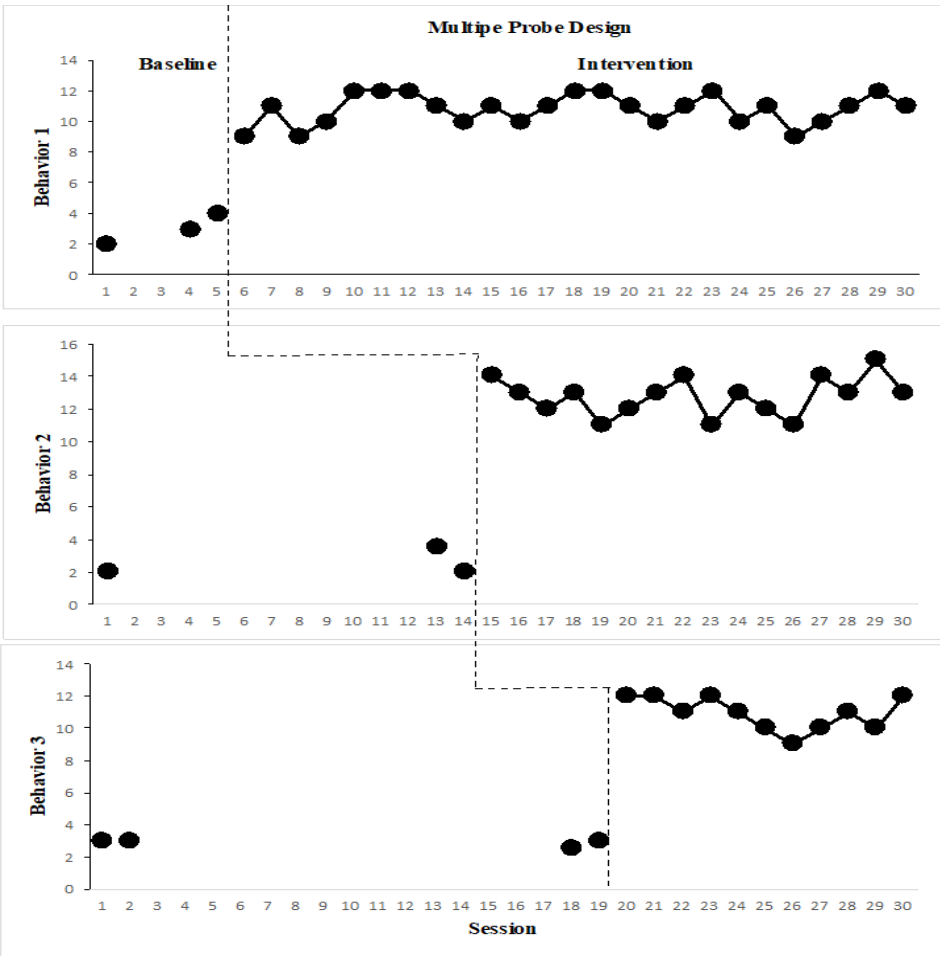

Figure 8. Multiple probe design.

Delayed Multiple Baseline Design.

In the delayed multiple baseline design, an initial baseline and intervention are initiated for one behavior, setting, or subject. Then, subsequent baselines for additional behaviors are introduced in a staggered or delayed fashion (Cooper et al., 2007).

Houston-Wilson, Dunn, van der Mars, & McCubbin, (1997) used a single-case delayed multiple baseline design across six student participants with developmental disabilities. The student participants' ages were 9 to 11 (5 boys and 1 girl). All students were classified as having mild mental retardation and were studying in a self-contained special education class. However, the students were regularly integrated into physical education, art, and music. Six typically developing peers served as peer tutors to the six student participants with mild mental retardation. The study aimed to investigate the effect of untrained and trained peer tutors on improving the motor performance of students with developmental disabilities in integrated physical education classes. The dependent variable in the study was the ability to perform critical elements of fundamental motor skills (the horizontal jump, catch, overhand throw, forehand strike, and sidearm strike) .

The researcher trained the peer tutors on three particular teaching skills: to use appropriate cues, feedback, and task analysis of motor skills. The peer tutors were given handouts that defined these teaching skills. They were also given a script of scenarios that communicated what appropriate instruction looked like. The results of this study indicated that the student participants who worked with untrained peer tutors showed no significant improvement in their mean percentage of motor skill appropriateness score compared to trained peer tutors. This was observed across student participants 1, 2, and 3. Trained peer tutors, however, assisted the student participants in improving their mean percentage of motor skill appropriateness score. This prediction was observed across participants 1 and 3. The results concluded that tutoring was efficient when the peer tutors were provided adequate training to ensure that their tutoring is focused on the target behaviors.

Generally, the advantage of multiple baseline design is that it does not require or depend on withdrawing treatment in order to prove that the behavior alteration is a function of the intervention. Therefore, there is no need to remove or temporarily pause the intervention for a while in order to demonstrate the functional relationship between an independent and dependent variable (Kazdin, 2011). This characteristic of multiple baseline design enables a researcher to assess the effect of interventions that cannot be withdrawn (Cooper et al., 2007), for example, like learning to read.

Visual Analysis in SCD

Cooper et al., (2007) stated six advantages of graphic display and visual analysis. First, plotting each measure of behavior on a diagram immediately after the observation period allows a researcher to evaluate a participant's performance continually on visual record. Second, graphic display enables the researcher to examine interesting variations in behavior as they happen. Third, diagrams identify statistical analyses of behavior alteration; however, visual analysis of graphs takes less time and can make the information easier to decipher. Additionally, visual analysis does not depend on statistical assumptions. Fourth, visual analysis is considered to be a conservative method for determining the significance of behavior alteration. Therefore, behavior alteration is considered statistically significant due to the data plotted on a graph that detected the range, variability, trends, and overlaps in the data. Fifth, visual analysis is effective in showing independent judgments and explanations of the behavior change. Sixth, visual analysis also illustrates and offers an effective source of feedback.

There are a number of types of graphs used to display behavioral data. Commonly used graphic displays include *line graphs*, *bar graphs*, *cumulative records*, s*emi-logarithmic charts*, and *scatter plots*. *Line graphs* are the model most commonly used (Cooper et al., 2007).

Visual Analysis of Data Considered Conservative in SCD

Visual analysis of single-case data is the fundamental method of disseminating the effects of an independent variable or a dependent variable (Kahng et al., 2010). Visual analysis is more effective than tests of statistical significance because statistical procedures are less conservative and are more likely to produce Type I errors than visual analysis of data (Poling & Fuqua, 1986) In special education, researchers are concerned with producing socially significant behavior changes, and furthermore, are not interested in knowing that a behavior change was statistically significant based on the result of an intervention. Baer (1977a) reported, “If a problem has been solved, you can *see* that; if you must test for statistical significance, you do not have a solution” (p. 171). Visual analysis is completely appropriate to recognize variables that can be described as robust, large, and reliable and that contribute to an efficient, strong technology of behavior change. Otherwise, tests of statistical analysis can reveal the lower potential correlation between the independent and dependent variables; thus, it might contribute to weak or unreliable variables in the technology (Cooper et al., 2007).

There is potential for two types of errors that can impact results: Type I error and Type II error. A Type I error takes place when the researcher rejects a true null hypothesis. If the null hypothesis is true, it should not be rejected. Type I errors are called *false positives* due to the researcher wrongly concluding that a relationship exists in the experiment. A Type II error occurs when the researcher fails to reject a false null hypothesis; if the null hypothesis is wrong, it is supposed to be rejected. Type II errors are called *false negatives* because the researcher wrongly concluded that no relationship exists (Johnson & Christensen, 2014). Therefore, in single-case design the researcher depends on visual analysis in order to confirm intervention effects, which results in low incidence of Type I error. However, there is a chance of *"increasing the commission of Type II errors" (*Cooper et al., 2007, p. 249). The researcher who depends on tests of statistical significance in order to assert intervention effects the most are committed to Type I errors more than the researcher using single-case design.

Visual analysis findings offer conservative conclusions regarding intervention effects. When compared to statistical inference, visual analysis yields a conservative result and impacts decision making (Camphell & Herzinger, 2010). For instance, Parsonson and Baer (1986) reported that research results appeared "*to imply that time-series analysis [a statistical technique] is usually less conservative than visual analysis"* (p. 159). A conservative decision results in a decrease in the probability of a Type I error (Camphell & Herzinger, 2010). Furthermore, it can be concluded that an effective functional relationship exists when in fact one does not. The conservative judgment that may result from visual analysis works as a "*filter"* in order to tease out weak intervention effects so that only strong intervention effects are applied. Therefore, the conservative affirmation of visual analysis corresponds with a differential tolerability for errors of inference: "*Type II errors are more acceptable than Type I errors"* (Camphell & Herzinger, 2010, p. 419). That is, ignoring small effects (i.e., Type II errors) is sometimes more desirable than wrongly concluding that effects are present when they in fact are not (i.e., Type I errors) (Camphell & Herzinger, 2010).

Overall, visual analysis has significant benefits, such as the ability of assessing an experimental effect conservatively, and finding variables that are primary and conducive to socially useful technology of behavior.

Conclusion

In special education research, it is quite difficult to have a sufficient number of participants to gather randomly from the population. Based on a report by the US Department of Education, in 2015, the number of children and youth ages 3–21 receiving special education services in the United States was about 13 percent of all public school students. Children who were classified as deaf or hard of hearing make up 1% of the general education population (US Department of Education, 2015). Implementing group experimental designs in special education with heterogeneous low-incidence populations can be hard. On the other hand, a SCD does not demand a large number of participants, and so researchers can readily implement interventions with students who are receiving special education services in order to improve academic performance.

Reference:

Alberto, P.A. & Troutman, A. C. (2013) *Applied Behavior Analysis for Teachers*. (9th ed). Boston, MA: Pearson Education.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*(1), 91-97.

Baer, D. M. (1977). Perhaps it would be better not to know everything. *Journal of Applied Behavior Analysis*, *10*(1), 167-172.

Belfiore, P. J., Basile, S. P., & Lee, D. L. (2008). Using a high probability command sequence to increase classroom compliance: The role of behavioral momentum. *Journal of Behavioral Education, 17*, 160-171.

Benedict, K. M., Rivera, M. C., & Antia, S. D. (2015). Instruction in metacognitive strategies t to increase deaf and hard-of-hearing students' reading comprehension. *Journal of Deaf Studies and Deaf Education, 20(1)*, 1-15

Cakiroglu, O. (2012). Single subject research: applications to special education. *British Journal of Special Education, 39*(1), 21-29.

Campbell J. M., Herzinger C. V. (2010). Statistics and single subject research methods. In Gast D. L. (Ed.), *Single subject research methodology in behavioral sciences* (pp. 417-453). New York, NY: Routledge.

Cannon, J. E., Guardino, C., Antia, S. D., & Luckner, J. L. (2016). Single-case design research: Building the evidence base in the field of education of deaf and hard of hearing students. *American Annals of the Deaf, 160*(5), 440–452.

Cooper, J., Heron, T., & Heward, W. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Dimling, L. M. (2010). Conceptually-based vocabulary intervention: second graders' development of vocabulary words. *American Annals of the Deaf, 155* (4), 425-448.

Hammond, D.& Gast, D.L.(2010) Descriptive analysis of single subject research designs:1983–2007. *Education and Training in Autism and Developmental Disabilities*, *45*(2), 187–202.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S. L. & Wolery, M. (2005) The use of single-subject research to identify evidence-based practices in special education. *Exceptional Children*, 71, 165–179.

Houston-Wilson, C., Dunn, J. M., Van Der Mars, H., & McCubbin, J. (1997). The effect of peer tutors on motor performance in integrated physical education classes. *Adapted Physical Activity Quarterly, 14(4)*, 298-313.

Johnston J.M, Pennypacker H.S. (2009) *Strategies and tactics of behavioral research* (3rd ed.) Hillsdale, NJ: Erlbaum.

Johnson, B., & Christensen, L. (2014). Educational research: Quantitative, qualitative, and mixed approaches (5th ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). Oxford, NY: Oxford University Press.

Kennedy, C. H. (2005) *Single-Case Designs for Educational Research*. Boston, MA: Pearson Education.

Kahng, S. W., Chung, K.-M., Gutshall, K., Pitts, S. C., Kao, J., & Girolami, K. (2010). Consistent Visual Analyses of Intrasubject Data. *Journal of Applied Behavior Analysis*. *43*(1), 35-45.

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 What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Mesibov, G. B., & Shea, V. (2011). Evidence-based practices and autism. *Autism*, *15*(1), 114-133.

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: randomization to the rescue. *Psychological Methods,15*, 2, 124-44.

Neuman, S. B. & McCormick, S. (1995) *Single-Subject Experimental Research: applications for literacy*. Newark, DE: International Reading Association.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. (2005).

Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, *2*, 137–148.

Paul, P., Wang, Y., & Williams, C. (2013). *Deaf students and the qualitative similarity hypothesis: Understanding language and literacy development*. Washington, DC: Gallaudet University Press.

Poling, A. D., & Fuqua, R. W. (1986). *Research methods in applied behavior analysis: Issues and advances*. New York, NY: Plenum Press.

Riley-Tillman, T. C., & Burns, M. K. (2009). *Evaluating educational interventions: Single-case design for measuring response to intervention*. New York, NY: Guilford Press. Plavnick, J. B., & Ferreri, S. J. (2013). Single-case experimental designs in educational research: A methodology for causal analyses in teaching and learning. *Educational Psychology Review*, *25*(4), 549-569.

Sealander, K. A. (2004) Single-subject experimental research: an overview for practitioners, in

K. deMarrais and S. D. Lapan (eds), *Foundations for Research*. Mahwah, NJ: Erlbaum.

Richards, S. B., Taylor, R. L., Ramasamy, R. & Richards, R. Y. (1999) *Single-Subject Research: applications in educational and clinical settings*. San Diego, CA: Singular.

Risley, T. (2005). Montrose M. Wolf (1935–2004). *Journal of Applied Behavior Analysis*, *38*,279–287.

Tankersley, M., Harjusola-Webb, S. & Landrum, T. J. (2008) Using single subject research to establish the evidence base of special education, *Intervention in School and Clinic*, 44, 83–90.

US Department of Education (2015) *Thirty-seventh Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act*. Washington, DC: US Department of Education [online at http://www2.ed.gov/about/reports/annual/osep/2015/pa