**International Journal of Advanced Research and Development**
**ISSN: 2455-4030**
**www.newresearchjournal.com/advanced**
**Volume 1; Issue 2; February 2016; Page No. 17-20**

# Quantitative approximations of truth: A technical synopsis of internal validity in group design and single-case design research

**Elias Clinton**
Department of Special Education, Black Hills State University, Spearfish, South Dakota 57799, United States of America.

**Abstract**
Internal validity is a scientific property that plays a critical role in the design, application, interpretation, and dissemination of experimental research. The strength of a study's internal validity is based on the extent to which researchers have demonstrated experimental control, controlled for confounding variables, and minimized systematic error. Strong internal validity allows researchers to analyze data and report the results of a study with confidence. The value of research and its impact on professional practice, policies, and society as a whole is unquestionable. Therefore, scholars/researchers must demonstrate a thorough understanding of threats to internal validity, and how these threats apply to specific research methodologies. This paper examines the definition of internal validity as it relates to group design and single-case design research. Furthermore, specific threats to internal validity are outlined and methods/experimental designs employed by quantitative researchers to minimize those threats are discussed.

**Keywords:** internal validity, group design research, single case design research, single subject design research, research methodology, quantitative research

## 1. Introduction
Researchers must design empirical studies with an emphasis on scientific properties such as internal validity in order to increase the value and relevance of research findings. This paper outlines the scientific property of internal validity (IV) as it relates to both group design and single-case design research. Additionally, this paper discusses threats to IV and provides examples of methods and experimental designs employed by quantitative researchers in order to minimize those threats.

## 2. Definition of internal validity
Quantitative researchers conduct studies by manipulating an independent variable and monitoring/recording any effects on a dependent variable. Internal Validity is based on the extent to which researchers have demonstrated experimental control, controlled for confounding variables, and verified the independent variable alone resulted in observed changes (if any) to the dependent variable (Tuckman & Harper, 2012) [1]. Internal validity is increased if a study has employed an appropriate research design and any necessary measures to ensure that no confounding variables resulted in the dependent variable changes. Further, IV is increased based on the extent to which the study minimized systematic error (Tuckman & Harper, 2012) [1]. Systematic error is an error that is introduced by an inaccuracy inherent in the system of measurement utilized in the research, and not by chance alone (Tuckman & Harper, 2012) [1]. Internal validity affects whether the findings of the research can be confidently accepted due to a design that demonstrated experimental control of the variables resulting in changes to the dependent variable (Tuckman and Harper, 2012) [1].

## 3. Threats to internal validity
The following provides examples of specific threats to internal validity that researchers must attempt to minimize when conducting experimental studies. Ambiguous Temporal Precedence refers to the lack of clarity regarding which variable is the cause and which variable is the effect (Kratochwill *et al*., 2010) [2]. Adaptation is a threat that occurs due to the participants' eventual acclimation to the novel stimuli of the study's conditions (Gast & Ledford, 2014) [3]. Attrition/Mortality is a threat to IV when analysis occurs only for participants that have participated in the entire duration of the study (Kratochwill *et al*., 2010) [2]. If participants drop out of an experimental study, potentially due to the study conditions, this can make it difficult for researchers to confidently determine if changes in the dependent variable were the result of the independent variable or particular characteristics of the remaining participants (Kratochwill *et al*., 2010; Tuckman & Harper, 2012) [2, 1]. Confounding Variables refers to variables (inner- and/or extra-experimental) that are not systematically manipulated by the researchers, yet may be responsible for observed changes in the dependent variable (Tuckman & Harper, 2012) [1]. Cyclical Variability may occur when the various conditions of a study are of equal length; therefore, any changes in the dependent variable may be the result of a variable that coincidentally occurs simultaneously with the study conditions, yet is not manipulated by the researchers (Gast & Ledford, 2014). Experimenter Bias is a threat to IV that refers to researcher behaviors based on anticipated outcomes that interfere or affect the interpretation and reporting of a study's results (Tuckman & Harper, 2012) [1]. The Hawthorne Effect refers to changes in the participants' behavior that occur as a direct result of being observed by the researchers (Gast & Ledford, 2014) [3]. History refers to events that may occur outside of the experimental conditions that may attribute to changes in the dependent variable (Kratochwill *et al*., 2010) [2]. Instrumentation refers to changes in the dependent variable being the result of the instrument or measurement system

used in the study (Gast & Ledford, 2014) [3]. That is, data collectors may develop increased familiarity with the instrument throughout the study, or change the criteria used to evaluate participants' performance while the study is occurring (Gast & Ledford, 2014) [3]. Maturation refers to the physical, emotional, or cognitive development that participants may experience during the course of a study (Kratochwill *et al*., 2010) [2]. Multiple-Treatment Interference refers to the changes in the dependent variable resulting from more than one treatment in a study (Gast & Ledford, 2014) [3]. Procedural Infidelity is a threat that occurs when the procedures of all conditions are not implemented as planned; therefore, procedural fidelity should be measured and reported in order to increase the confidence of findings (Billingsley, White, & Munson, 1980) [4]. Regression toward the Mean refers to the phenomenon that occurs when participants are chosen due to outlying scores (Kratochwill *et al*., 2010) [2]. If the outlying score is not a valid measurement of a participant's skill level, his/her score may improve on subsequent measures due to a regression to the mean rather than the manipulation of the independent variable. Selection Bias refers to observed differences being attributed to idiosyncratic status variables of participants instead of manipulations to the independent variable (Tuckman & Harper, 2012) [1]. Testing is a threat when using repeated measures in a study because increased performance might be related to participants' familiarity with the assessment items (i.e., a facilitative effect) rather than true skill development (Gast & Ledford, 2014; Kratochwill *et al*., 2010) [3, 2]. Repeated testing may also have an inhibitive effect (i.e., fatigue/reduced response effort) if the participant is tested repeatedly but not accessing reinforcement (Gast & Ledford, 2014) [3].

**4. Internal validity and group design research**

Group design experimental research typically involves comparing the performance of at least two different groups of participants (Tuckman & Harper, 2012) [1]. A typical group design study involves comparing a control group (i.e., participants that do not receive treatment) to an experimental group (i.e., participants that receive treatment). Group researchers determine which quantitative techniques (e.g., measures of central tendency, analysis of variance, correlation and regression analyses, nonparametric statistical tests) are needed to interpret the data based on their research questions, and use statistical analysis to compare groups' performances and determine the probability that chance variations produced any observed differences (Tuckman & Harper, 2012).

Researchers that employ quantitative group experimental designs attempt to control for threats to internal validity in a variety of ways. Researchers can increase IV by describing all research conditions, participants, measures, and procedures with sufficient detail that would allow for replication by an independent researcher (Gast & Ledford, 2014) [3]. Threats to internal validity are also controlled for using multiple measures throughout the course of a study. For example, pre-/post-test designs are common, but studies that include repeated measures allow researchers to conduct more in-depth statistical analyses (Gast & Ledford, 2014) [3]. Researchers that utilize group designs must also confirm that data collection methods are used in a manner appropriate for the research design and questions (Gast & Ledford, 2014) [3]. Ensuring that researchers/data collectors are equally unaware or aware of participant characteristics and the study's research hypotheses decreases bias and consequently increases IV. Data collectors who are blind to study details, except behavioral definitions, are preferred in order to control for biased recording. Thus, group design researchers may employ a blind or double-blind system to ensure that data collection is not influenced by bias or researcher expectations (Tuckman & Harper, 2012) [1]. Procedural fidelity and reliability checks of all variables can also increase a study's IV (Gast & Ledford, 2014) [3]. Finally, IV is strengthened when participants are randomly assigned or counterbalanced to groups/conditions (Tuckman & Harper, 2012) [1]. Random assignment is critical because it reduces the potential that one group performed significantly higher than the other group based on status variables (e.g., age, gender, socio-economic status, level of cognitive functioning, learning history) (Tuckman & Harper, 2012) [1]. If participants are randomly assigned, researchers can be more confident that changes in the dependent variable were likely the effect of the independent variable and not the group's idiosyncratic characteristics. Essentially, random assignment is presumed to more evenly distribute the baseline variance and allows the researchers to assume that any differences in baseline were due to chance and not to systematic assignment. Researchers should report the number of participants that withdrew during the course of the study and from which condition (Tuckman & Harper, 2012) [1]. Collecting and reporting attrition data indicates whether or not participant withdrawal was equal across groups/conditions (Gast & Ledford, 2014; Tuckman & Harper, 2012) [3, 1]. If attrition was higher for one group, researchers cannot say with confidence that the groups were equivocal (Tuckman & Harper, 2012) [1].

**5. Internal Validity and Single-Case Design Research**

Single-case design (SCD) research is a methodology most often used in applied fields of human behavior (e.g., healthcare, psychology, education) in which individual units serve as their own control (Richards *et al*., 2000) [6]. That is, individual cases are the unit of intervention and analysis (Kratochwill *et al*., 2010). The SCD experimental principle of each unit serving as its own control is referred to as baseline logic (Cooper, Heron, & Heward, 2007; Gast & Ledford, 2014; Richards, Taylor, Ramasamy, & Richards, 2000) [5, 3, 6]. SCD is most often employed when the research question targets differences in an individual's data, and data are measured frequently as well as across or within conditions/phases or varying levels of the independent variable. Comparatively speaking, SCD is sensitive to individual unit differences, and group designs are sensitive to the average data of a group (Richards *et al*., 2000) [6].

The dependent variable in SCD is measured frequently and across or within conditions/phases or varying levels of the independent variable (Gast & Ledford, 2014) [3]. Baseline is the first phase in most SCD studies. Baseline involves measuring a specified dimension of a target behavior/ dependent variable. Baseline does not necessarily mean that "nothing" is occurring in the participant (s) environment, but rather represents a "business-as-usual" measurement (Richards *et al*., 2000) [6]. Following baseline, researchers introduce intervention in the next condition and examine the effects on the dependent variable. During conditions,

researchers typically examine trend, level, and variability of the data (Kratochwill *et al*., 2010) [2].

Similarly to group design studies, single case research designs must control for threats to IV to increase the confidence of findings. Single case research can address threats to IV through replication of effect within a study (Kratochwill *et al*., 2010) [2]. Replication of effect is important for increasing IV, and for demonstrating experimental control. Horner *et al.* (2005) [7] provided research criterion for replication that involved three replications of effect at three different points in time. In other words, covariance of predicted changes in the dependent variable and manipulations of the independent variable at three different points in time, or across three different cases, is indicative of minimized threats to IV and strong experimental control (Horner *et al*., 2005; Kratochwill *et al*., 2010) [7, 2]. However, it is possible for a study to have strong IV, yet not demonstrate three demonstrations of effect at three different points in time. That is, changes in the dependent variable may not have been observed, yet a study's IV may still have been strong as long as the study controlled for extraneous variables and appropriately measured and addressed the research questions (Richards *et al*., 2000) [6]. The aforementioned scenario would simply indicate that the intervention was not effective for a particular individual in a particular context.

Specific experimental designs inherent to SCD allow researchers to demonstrate replications of effect and increase the study's IV. For example, the ABAB or withdrawal design (i.e., baseline phase, intervention phase, return to baseline, intervention phase) allows for three replications of effect at three different points in time (Kratochwill *et al*., 2010). This design can demonstrate a strong causal relationship between predicted changes in the dependent variable and strategic manipulations of the independent variable (Gast & Ledford, 2014; Kratochwill *et al*., 2010) [3, 2]. The ABAB design minimizes the threats of history and maturation because the researcher has demonstrated that changes in the dependent variable only occur with intended manipulations of the independent variable (Richards *et al*., 2000) [6]. See Figure 1 for an example of an ABAB design that demonstrates three replications of effect at three different points in time (i.e., a functional relation) via a graphed set of data.

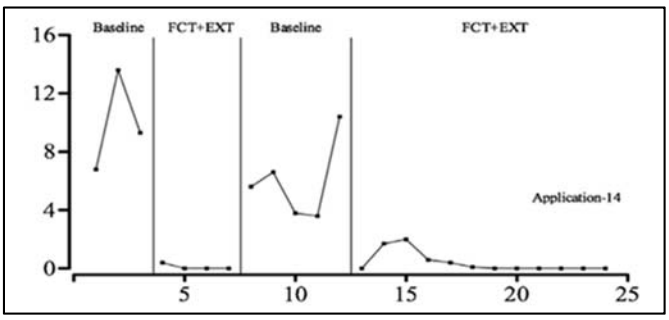


**Fig 1:** Example of an ABAB single-case research design.

From: Rooker, G. W., Jessel, J., Kurtz, P. F., & Hagopian, L. P. (2013). Functional communication training with and without alternative reinforcement and punishment: An analysis of 58 applications. Journal of Applied Behavior Analysis, 46(4), 708-722 [8].

Another SCD design that can demonstrate a functional relation and control for threats to IV is the multiple baseline design (MBD) (Kratochwill *et al*., 2010) [2]. A MBD is essentially a series of A-B designs (baseline – treatment) that can be staggered across participants, behaviors, or settings (Richards *et al*., 2000) [6]. Typically a MBD involves repeated measures concurrently across at least two baselines (Kratochwill *et al*., 2010) [2]. To strengthen IV, intervention is implemented only after responding is stable for all baselines (Gast & Ledford, 2014) [3]. The independent variable can be implemented across all baselines simultaneously; however, introduction of intervention can be staggered timewise in order to demonstrate at least three demonstrations of effect at least three different points in time (Horner *et al*., 20005) [7]. The threats of history, maturation, and testing are controlled for using the staggered introduction of the independent variable across baselines (Gast & Ledford, 2014) [3]. See Figure 2 for an example of a MBD.

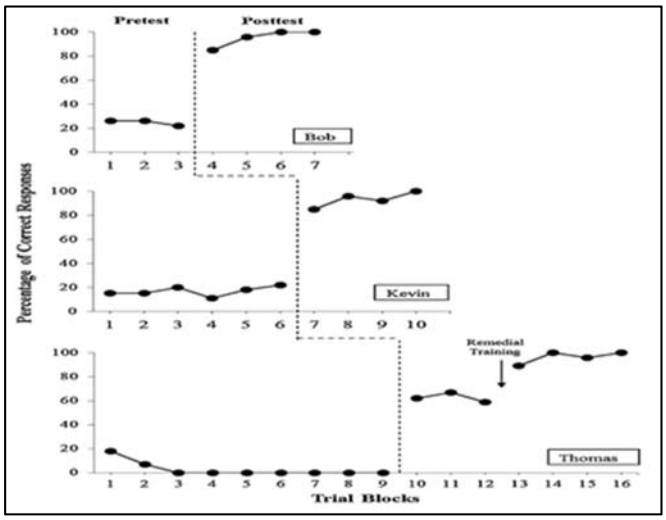


**Fig 2:** Example of a multiple baseline design.

From: De Souza, A. A., & Rehfeldt, R. A. (2013). Effects of dictation-taking and match-to-sample training on listing and spelling responses in adults with intellectual disabilities. Journal of Applied Behavior Analysis, 46(4), 792-804 [9].

Another common SCD design, that is used to compare two or more interventions, is referred to as the alternating treatment or multi-element design (Kratochwill *et al*., 2010) [2]. In most cases the alternating treatment design (ATD) involves rapidly alternating at least two interventions in a counterbalanced or random order of presentation and examining the effects of each intervention on the dependent variable (Richards *et al*., 2000) [6]. ATD designs may or may not include a baseline condition or no treatment condition within the study. Some scholars have posited that the alternating treatments design is capable of determining a functional relation, yet is weaker than other SCD designs (Alberto & Troutman, 2012) [10]. Other researchers contend that ATD allow for prediction, verification, and replication (i.e., baseline logic) because each data point serves as a predictor for future behavior under the same condition, each data point serves as a verification of previous performance predictions under the same condition, and each data point replicates the differential effects produced by the other treatments (Richards *et al*., 2000) [6].

The IV of a study can be increased through the use of an ATD if practice/testing effects are a concern (Richards *et al*., 2000) [6]. That is, the counterbalanced or random presentation of treatments minimizes the threat of sequencing effects. ATD minimizes the threat of maturation and history based on the relatively short time frame required to implement the design (Gast, & Ledford, 2014) [3]. Procedural drift is minimized by training data collectors sufficiently and retraining as needed and detected with procedural fidelity data collection on a frequent basis (Gast & Ledford, 2014) [3]. Multi-Treatment interference is a possible threat for ATD during the alternating condition; however, researchers often carry out a superior treatment only phase which minimizes any threats of a carryover effect (Gast, & Ledford, 2014) [3]. See Figure 3 for an example of graphed data from an ATD.

Other threats to IV such as selection effects and regression to the mean are not typically a concern based on the nature of the research questions inherent in SCD; however, these threats may need to be controlled for under certain conditions (Kratochwill *et al*., 2010) [2]. Ambiguous temporal precedence is controlled for through phase repetition and replication of effect using any of the various designs previously discussed in order to demonstrate active manipulation of the independent variable (Kratochwill *et al*., 2010; Richards *et al*., 2000) [2, 6]. The threat of attrition is also problematic in SCD; therefore, researchers should include an adequate number of participants so that in the case of participant withdrawal, enough data can be collected to conduct meaningful analysis. Instrumentation and observer drift are often minimized through the use of procedural fidelity checks and interobserver agreement in at least 20% of all conditions with all participants (Gast & Ledford, 2014; Kratochwill *et al*., 2010) [3, 2].

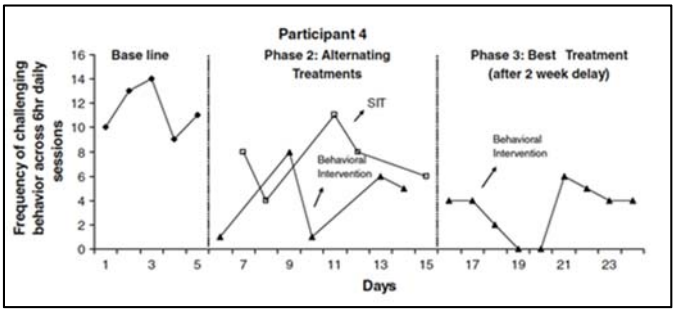


**Fig 2:** Example of a graph from an alternating treatment design.

From: Devlin, S., Healy, O., Leader, G., & Hughes, B. M. (2011). Comparison of behavioral intervention and sensory-integration therapy in the treatment of challenging behavior. Journal of Autism and Developmental Disorders, 41(10), 1303-1320 [11].

## 6. Conclusions

Controlling for threats to IV is essential for researchers to convey research results with confidence. Valid research is critical to the development of evidence-based practices; therefore, it is paramount that researchers choose research designs that effectively minimize threats to IV in order to provide scientific evidence needed to enhance their respective fields of study.

## References


1. Tuckman B.W, Harper B.E. Conducting educational research. Plymouth, MD: Rowman & Littlefield Publishers, 2012.
2. Kratochwill T.R, Hitchcock J, Horner R.H, Levin J.R, Odom S.L, Rindskopf D.M, Shadish W.R. Single-case designs technical documentation. What Works Clearinghouse, 2010.
3. Gast D.L, Ledford J.R. (Eds.). Single Case Research Methodology: Applications in Special Education and Behavioral Sciences, 2e: Applications in Special Education and Behavioral Sciences. New York, NY: Routledge, 2014.
4. Billingsley F.F, White O.R, Munson R. Procedural reliability: A rationale and an example. Behavioral Assessment 1980; 2(2):2.
5. Cooper J.O, Heron T.E, Heward W.L. Applied Behavior Analysis (2nd ed.). Columbus, OH: Prentice Hall, 2007.
6. Richards S.B, Taylor R.L, Ramasamy R, Richards R.Y. Single subject research. San Diego, CA: Singular Publishing Group, 2000.
7. Horner R.H, Carr E.G, Halle J, McGee, G, Odom S, Wolery M. The use of single-subject research to identify evidence-based practice in special education. Exceptional Children 2005; 71(2):165-179.
8. Rooker G.W, Jessel J, Kurtz P.F, Hagopian L.P. Functional communication training with and without alternative reinforcement and punishment: An analysis of 58 applications, Journal of Applied Behavior Analysis. 2013; 46(4):708-722.
9. De Souza A.A, Rehfeldt R.A. Effects of dictation-taking and match-to-sample training on listing and spelling responses in adults with intellectual disabilities, Journal of Applied Behavior Analysis. 2013; 46(4):792-804.
10. Alberto P.A, Troutman A.C. Applied behavior analysis for teachers, 9e. Upper Saddle River, NJ: Pearson Higher Ed, 2012.
11. Devlin S, Healy O, Leader G, Hughes B.M. Comparison of behavioral intervention and sensory-integration therapy in the treatment of challenging behavior, Journal of Autism and Developmental Disorders. 2011; 41(10):1303.