

Spencer, J. & Winkler, C. (2025). Single-Subject Research Design: A rigorous, responsive, and accessible approach to assessment and evaluation in higher education. *Intersection: A Journal at the Intersection of Assessment and Learning*, 6(3), 177-199.

Single-Subject Research Design: A Rigorous, Responsive, and Accessible Approach to Assessment and Evaluation in Higher Education

Jordan Spencer, Ph.D. & Christa E. Winkler, Ph.D.

Author Note

Jordan Spencer, <https://orcid.org/0000-0001-5066-9616>

Christa E. Winkler, <https://orcid.org/0000-0002-1700-5444>

We have no conflicts of interest to disclose.

Intersection: A Journal at the Intersection of Assessment and Learning

Vol 6, Issue 3

Abstract: Assessment, evaluation, and research in higher education serve as a means to evaluate and document evidence of educational effectiveness. However, many existing approaches to assessment and evaluation lack rigor in their methodological design, responsiveness to changing student needs, and accessibility to researchers and practitioners with varying degrees of methodological training. This paper offers an illustrative example of single-subject research design (SSRD) as a means to evaluate the effectiveness of a unique behavioral program developed by the staff of a post-secondary transition program on a university campus. Through this example, we highlight SSRD as a promising option for rigorous, responsible, and accessible assessment to improve data-driven decision-making in higher education.

Keywords: *Single-subject research design, assessment, evidence-based practice, data-driven decision making*

Introduction

As student enrollment in higher education continues to evolve, college administrators must grapple with key questions surrounding retention, persistence, and graduation (Adelman, 1999; Braxton & Hirschy, 2005; Marsh, 2014; Seidman, 2005) for an increasingly diversified student body. Some have even questioned the value of a college degree in our current society; for example, The Wall Street Journal recently published a telling essay titled “Why Americans have lost faith in the value of college” (Belkin, 2024). Such criticism highlights an urgent need for colleges and universities to improve outcomes for the changing student populations that they serve—and to also provide evidence of those improved outcomes.

With calls for assessing, and ultimately improving, student outcomes in higher education, conversations have centered around the need to create cultures of evidence to improve institutional accountability (Hersh & Keeling, 2013). Consideration of innovative assessment, evaluation, and research (AER) approaches can foster improved data-driven decision-making (DDDM) practices in higher education. In fact, Hong and Moloney (2020) argue that now is the time for the field of higher

education to deeply reflect, deconstruct, and intentionally rebuild its data collection and analysis practices to better align with the world in which we live and the learners we serve. Looking to innovative approaches that can supplement current AER practices can empower higher education researchers and practitioners to make data-driven decisions regarding the allocation of time, money, and resources to best serve students.

Need for Evidence-Based Practice in Higher Education

While there are many institutions answering the call to improve their AER culture and practices, most have not achieved systematic or systemic change (Hersh & Keeling, 2013). Additionally, many institutions claim to be utilizing assessment data to make decisions but are superficially answering external calls for accountability by spending hundreds of thousands of dollars, annually, to cover the cost of assessments that are intended to demonstrate institutional effectiveness (Cooper & Terrell, 2013; Cox et al., 2017). In fact, in a survey conducted by Jaschik and Lederman (2020), almost a third of chief academic officers thought assessment was intended for the appeasement of politicians and accreditors as opposed to improving teaching and learning. Additionally, a fifth of respondents did not agree that systems of assessment led to actual improvements in their universities (Jaschik & Lederman, 2020). Similarly, Fuller et al. (2015) found that higher education assessment leaders noted disconnects between the narrative surrounding assessment on their campuses – with an emphasis on improvement of student learning – and the actual assessment practices implemented on their campuses.

Collectively, these findings call higher education practitioners and researchers to expand the identification and utilization of evidence-based practices to improve student outcomes. What Works Clearinghouse (WWC) was established in response to the No Child Left Behind Act (NCLB, 2002) as an initiative of the Institute of Education Sciences (IES) at the U.S. Department of Education. WWC serves as a “central and trusted source of scientific evidence for what works in education” and aims to improve the utilization of “rigorous and relevant research, evaluation, and statistics to improve [the] nation's education system” (IES, n.d., para. 1). Although the WWC initially focused on K-12 education, it has since expanded to include evidence of what works in higher education. However, evidence-based programs, policies, and practices in higher education have not kept pace with those available to K-12 educators and administrators. To date, only 17 interventions associated with postsecondary education appear in the WWC database as compared to a total of 140 interventions associated with K-12 education that meet standards of rigor established by WWC. Compared to their K-12 education counterparts, higher education researchers and practitioners must contend with a scarce evidence base regarding what works for their students.

Need for Rigorous, Responsive, and Accessible Approaches to AER in Higher Education

There are several notable challenges faced by higher education researchers and practitioners, which contribute to the limited availability of robust evidence to guide practice. Although the importance of AER is recognized in professional standards for higher education and student affairs administrators (e.g., ACPA, NASPA, AERA/NCME), many have noted concerns that current skillsets and practices in higher education fail to meet expected standards in assessment, evaluation, and research competencies. Blimling (2001) stated, “If scholarship and practice in student affairs sometimes seem segmented, confused, and conflicted, the reason may be that they are” (p. 381). Subsequently, Sriram

and Oster (2012) found that 64% of professionals reported their engagement with AER was less than they desired, citing lack of time and access as the largest obstacles.

Other research in higher education further supports the inaccessibility of rigorous methods in everyday practice as a barrier to AER methods. For example, Mayhew et al. (2016) indicated that the most ideal way to measure student success is with designs that establish causality. Randomly assigning students to either a “treatment” or “control” group is preferred, as the effect of the treatment can be isolated, established, and a causal conclusion can be made. However, causal designs are typically unrealistic and/or unethical in college settings (Mayhew et al., 2016). Furthermore, most assessment approaches considered to be “rigorous” require specialized training in quantitative analytics—a sometimes intimidating or inaccessible goal for some practitioners (Sriram, 2014).

Regardless of the obstacles, AER skills play an increasingly important role in higher education practice (Hoffman & Bresciani, 2010; Young, 2001). Otherwise, higher education practice may become “simply random activity, bound by tradition and convention, maybe helpful, maybe not, probably suiting some students, almost certainly leaving others out” (Carpenter, 2001, p. 305). As such, there is a need for higher education to begin developing better connections between research and practice by helping practitioners see the value of AER and providing access to rigorous assessment methods that are accessible, practical, and can be readily applied in everyday settings. Given these challenges, higher education researchers and practitioners must begin considering alternative ways to develop evidence of effective educational practices that are rigorous in methodological design, responsive to changing student needs, and accessible to researchers and practitioners alike.

Single Subject Research Design (SSRD)

One option for higher education to consider is single-subject research design (SSRD). With a similar structure to correlational research, participant(s) in SSRDs act as their own control and do not require a comparison group (Kratochwill et al., 2010). Additionally, this design can show causality with a small number of participants (Smith & Little, 2018), addressing higher education’s difficulties in creating experimental and control groups. Moreover, studying a smaller number of participants reduces the cost of studies and the strain on resources that are typically needed to run a large-scale experiment, without compromising rigor of the design and associated findings.

While many studies dismiss small *n* research (Smith & Little, 2018), the success SSRD has had in creating DDDM and effective AER practices should not be in doubt, as fields like psychology (Smith & Little, 2018), Applied Behavior Analysis (Roane et al., 2016), and vision science (Ross, 2009) have found success in its use. Yet, the most promising evidence of single-subject working in higher education comes from Organizational Behavior Management (OBM; Erath et al., 2021). OBM, a subspecialty of Applied Behavior Analysis, focuses on changing and assessing the workplace by focusing on and improving employee performance and workplace culture (Behavior Analyst Certification Board, 2024); a focus and approach that seems to best fit the structural makeup of higher education institutions.

Across Applied Behavior Analysis and its subspecialties, however, there are numerous examples of SSRD applications in college settings. For example, researchers have focused on such topics as the

provision of effective feedback (Tittelbach et al., 2007; Wilk & Redmon, 1998) and identifying components of a college program that was effective in increasing social engagement in college students with autism (Fairchild et al., 2020). These, along with many other studies, provide multiple examples of how SSRD can be applied to assess and measure outcomes of both students and employees or faculty of higher education, both at program and individual participant levels.

Purpose

This paper illustrates the utility of single-subject research design (SSRD) as a robust method for higher education assessment. We describe the process of applying SSRD to evaluate S.A.F.E. (Student, Adult, Friend, Employee), an initiative implemented in one postsecondary transition program in a university setting. The S.A.F.E. program aimed to develop competencies related to academics, independent living, socialization, and employment for students with disabilities on a college campus. In particular, SSRD was used to systematically evaluate the effectiveness of teaching rules associated with the program. Programs such as the one in which S.A.F.E. was implemented are increasingly common in higher education settings, which highlights the importance for more research on successful programming. Yet, while the current data set reflects work with a population that has some form of disability, the reader should be aware that the intention of this paper is to showcase the use of single subject design within a real-world higher education setting. The evaluation of the current program was implemented within a student-affairs led program with typical resources and personnel associated with student affairs. Ultimately, the applicability and utility of SSRD can be expanded to countless other programs and offerings in higher education.

The SSRD application on which this paper focuses was developed and implemented by educational practitioners. In this paper we highlight the accessibility of SSRD – a rigorous, experimental research approach that relies heavily on visual, as opposed to statistical, analysis – to higher education professionals with varying levels of research skills and training. It is our hope that this paper serves as an illustration of the potential for researchers and practitioners, alike, to adopt robust approaches to evaluating and informing higher education practice using SSRD.

Methods

Data Source

Data used in this study were collected during the implementation of a post-secondary transition program referred to as the S.A.F.E. program. S.A.F.E. was created to support the development of competencies in academics, independent living, socialization, and employment (HEOA, 2008) for students with intellectual and developmental disabilities on a college campus. The data obtained for this study included data collected by S.A.F.E. program staff from fall 2019 to spring 2020 at a land grant institution located in the southeastern region of the United States. Program participants included 25 students with intellectual and/or developmental disabilities with ages ranging from 18-26 years (see Table 1 for sample demographics).

Table 1.*S.A.F.E. student participant demographics*

		Program (N = 25)	
		<i>n</i>	%
Racial/Ethnic Identity			
	White	18	72
	Black/African American	6	24
	Another racial identity	1	4
Gender Identity			
	Man	19	76
	Woman	6	24
Classification			
	Freshman	6	24
	Sophomore	7	28
	Junior	7	28
	Senior	5	20

Research Design

SSRD involves the repeated, systematic measurement of a dependent variable before (e.g., baseline), during (e.g., intervention) and after (e.g., post-intervention) the active manipulation of an independent variable (Kratochwill et al., 2010). In this example of SSRD, a reversal design was used where the dependent variable was the number of S.A.F.E. referrals submitted by staff, and the independent variable was the approach used to motivate student compliance with S.A.F.E. rules. In other words, this study sought to address an important question of practice—which approach resulted in fewer S.A.F.E. referrals: simply teaching the rules, providing rewards for following the rules, or a combination of both approaches. As should be done in all SSRD studies, the costs and benefits of using this approach to evaluate the alteration of student motivations were weighed with consideration for the well-being of participants involved.

Independent Variable

The S.A.F.E. program was comprised of two primary components: (1) teaching S.A.F.E. “rules” for student behavior, and (2) providing rewards for following rules. S.A.F.E. rules characterized behaviors of a successful student, adult, friend, and employee. Rules were taught consistently and systematically at the beginning of every class meeting, and they were posted in student rooms and a shared meeting space.

Dependent Variable

The assessment of S.A.F.E. program effectiveness relied on tracking student referrals that were submitted by full-time staff and student workers who were assigned to work more on a one-to-one

basis with students in the program. Students with behaviors that violated rules received S.A.F.E. referrals (see Table 2 for examples), which were tracked electronically; at certain points in the program, students who had 2 or fewer referrals per week received rewards. While this approach is more clinical it can easily be generalized to the larger institution as higher education houses multiple programs where students are provided with criteria that they must meet. In the classroom, many colleges set criteria for how many D's and F's one can receive before academic probation or dismissal from the university. The dean of students may set criteria that a student must adhere to with regards to conduct in order to avoid suspension or dismissal. Fraternities and sororities have criteria that students must adhere to in order to retain membership.

Table 2.

S.A.F.E. referral categories and corresponding behaviors

Referral Category	Behaviors
Student	Late to class, Sleeping in class, Verbal disrespect to staff, Late assignment, Audit course check-in form (did not submit), Audit course check-in form (negative feedback on form), Phone usage in class, Not following staff direction, Other
Adult	Failed room check, Breaking curfew, Late to meeting, Violating guest rules, Breaking roommate agreement, Lost key/locked out of room, Other
Friend	Non-aggressive argument, Inappropriate interaction with stranger, Name-calling, Inappropriate phone/text/GroupMe interaction, Other
Employee	Not following supervisor's instruction, Phone usage at work, Not wearing uniform/dress code violation, Negative report from supervisor, Inappropriate conversation, Inappropriate behavior, Late to work, Other

Design Approach

SSRD allows researchers the flexibility of choosing a design that fits the research question and the environment in which it is being measured. After data were obtained, the pattern of S.A.F.E. implementation was analyzed to determine how the independent variable was introduced and manipulated to determine the design of best fit, which could include common designs like the multiple baseline design, alternating treatments design, changing criterion design, and the reversal design. Upon reviewing the implementation of the S.A.F.E. program it was observed that staff systematically introduced and removed components of the independent variable, while consistently measuring the dependent variable.

More specifically, designated staff started the school year by training the staff on submitting referrals but did not introduce the independent variable (S.A.F.E) at this time. In SSRD this would be referred to as condition "A" or baseline, which acts as control for the participant. Staff then introduced a component of the independent variable by teaching S.A.F.E. to the students, hanging posters, and reviewing the program in all classes. This phase would be considered condition "B" or a treatment

condition in SSRD. From there staff systematically alternated between the “A” and “B” conditions throughout the fall, ending in treatment condition “B” before winter break. Upon returning, staff again started with the “A” condition before they introduced an incentive component that rewarded students for receiving two or less referrals in a week, which occurred in conjunction with the continuation of reviewing the rules of S.A.F.E and hanging of posters seen in condition “B.” After this staff returned to the “A” condition before discontinuing the program.

The pattern in which staff manipulated the independent variable best matched that of a reversal design, which must minimally include an initial baseline where the independent variable is absent, an intervention condition that introduces the independent variable, and a return to baseline condition that withdraws the independent variable (Cooper et al, 2020). For this specific data set, the reversal design would be labeled as an ABABA(B+C)A reversal design (see Table 3 and Figure 1).

Table 3.

Components of independent variable present for S.A.F.E. treatment conditions

Phase	A Baseline	B Treatment	B+C Treatment
Component s of IV Present	None	Teach S.A.F.E. rules Hang posters in foyer Hang flyer in student rooms	Teach S.A.F.E. rules Hang posters in foyer Hang flyer in student rooms Reward for 2 or less referrals

Figure 1.

S.A.F.E. treatment conditions timeline for single-subject reversal design



Analytical Plan

Visual analysis and computation of effect sizes were used to determine if one or more approaches utilized to motivate compliance with the S.A.F.E. rules (the independent variable; teaching rules and/or providing rewards) influences the number of referrals S.A.F.E. rule violations (the dependent variable).

Visual Analysis

Visual analysis involves the assessment of six variables: (1) level, (2) trend, (3) variability, (4) immediacy of change, (5) degree of overlap, and (6) consistency of data patterns (Horner et al., 2012; Kratochwill, et al., 2010). A summary of each variable, and associated interpretations are described in Table 4. To highlight the process of conducting visual analysis across each variable, a hypothetical dataset (see Appendix) was used for each variable. The dependent variable, “numbers correct,” represents number of correct test responses – for example, in a math course – both with and without the presence of the independent variable, such as a specific teaching strategy for math. The relationship between the teaching strategy and number of correct responses would be measured across time, with each repeated assessment labeled as a trial on the graph.

Table 4.

SSRD visual analysis variables and associated interpretations

Variable	Definition	Interpretation (Horner et al., 2012)
Level	Central measure of tendency (e.g., mean, median) of all data points within a phase	Retrieved by comparing the level data from the first phase to the level data in the second phase; process is repeated with each new phase
Trend	Slope of the line that best fits the data within each phase	Can be described as increasing, decreasing, or zero trend The closer data within each phase are to the trend line, the more stable The greater the difference in the slope of trend lines is between adjacent phases, the more confidence that there are differences between phases The closer the data points are to the trend line, the less variability exists within the phase
Variability	Deviation (range) of scores around the trend line	Changes in variability between phases can indicate a treatment effect, even if no changes in level and trend are observed
Immediacy of change	Immediacy of change in a data pattern following the manipulation of the independent variable	Retrieved by comparing the last 3-5 data points from one phase and the first 3-5 data points from the next phase The greater the immediacy of effect, the more likely a change is associated with the manipulation of the independent variable
Degree of overlap	Proportion of data points in a treatment phase that overlaps with the data points from the previous baseline phase	Low overlap suggests a larger effect

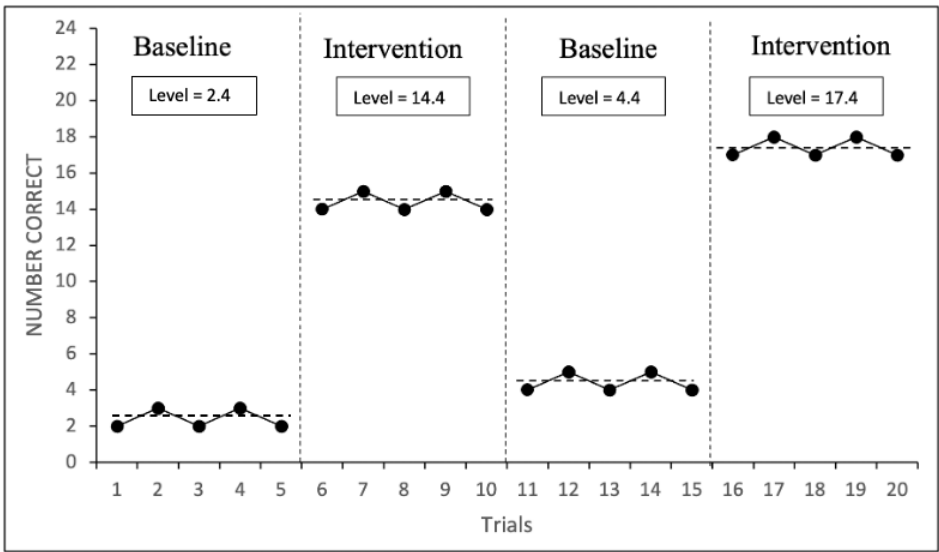
Consistency of data patterns	Extent to which there is consistency in data patterns on all phases within the same condition (e.g., all “A” conditions, all “B” conditions)	Phases generally need to demonstrate an effect at three different points in time to consider the independent and dependent variables functionally related The consistency, replication, and magnitude of the effect determine the level of confidence in an effect between the independent and dependent variable
-------------------------------------	--	--

Level

Analysis of the level requires analysis of the mean score for the data within a phase. This portion of visual analysis, as exemplified in the hypothetical dataset below (Figure 2), allows for the comparison of baseline and treatment phases across time to see if there was an increase or decrease in the mean of the dependent variable (labeled “number correct”) after the independent variable (teaching strategy) was introduced (Kratochwill, et al., 2010).

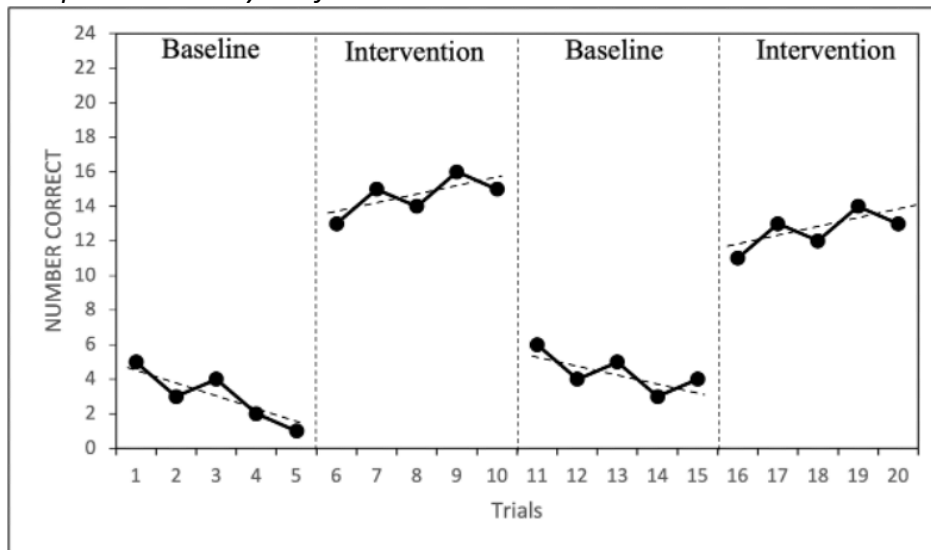
Figure 2.

Sample visual analysis of level in SSRD

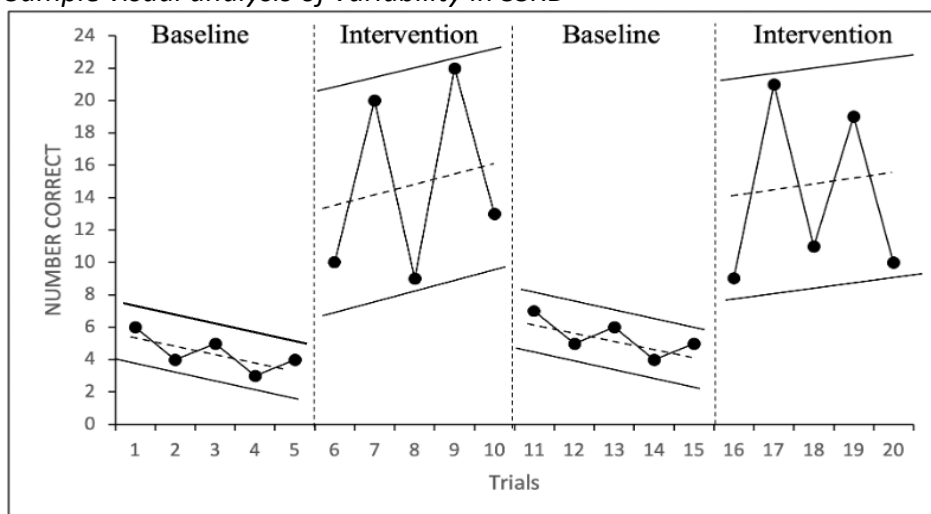


Trend

Visual analysis of the trend includes finding the slope of the best-fitting straight line for data within a phase, which can be described as increasing, decreasing, or a zero trend (Kratochwill, et al., 2010). The analysis of trend supplements the measure of level, which often cannot fully describe the change in the dependent variable. As highlighted below in the hypothetical data (Figure 3), just reporting the average of baseline or treatment would not be a complete description of the increase or decrease observed. Additionally, if data is extremely variable, reporting on trend can help to decipher if a positive or negative change has occurred.

Figure 3.*Sample visual analysis of trend in SSRD***Variability**

Variability analyzes the range of data around the best-fitting straight line determined by trend (Kratochwill, et al., 2010). Reporting on variability can reflect the strength of the relationship between the introduction of an independent variable and the change in a dependent variable; however, there are times that variability may also represent other variables that are present and impacting results. To provide an example, the hypothetical data set below (Figure 4) shows baseline to have low variability with a decreasing trend, while intervention shows high variability and a positive trend. From this analysis, one may determine that while a change occurred, there may need to be adjustments made to the independent variable in order to strengthen its relationship to the dependent variable.

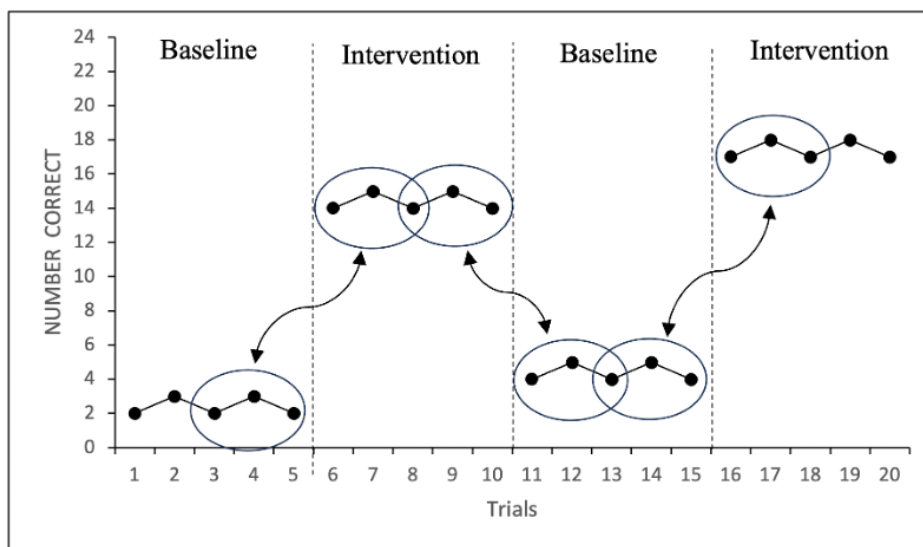
Figure 4.*Sample visual analysis of variability in SSRD*

Immediacy of Change

Immediacy of change is examined by analyzing the change in level between the last three data points in one phase and the first three data points of the next (Kratochwill, et al., 2010). To highlight the insight that this process can provide, the hypothetical data below (Figure 5) shows that the change in the dependent variable occurs immediately after introducing the independent variable with clear separation and no overlap in the adjacent phase. If the same effect is replicated, immediacy of change can provide further support that the introduction of the independent variable was the cause for the change in the dependent variable and that its effect was a strong one.

Figure 5.

Sample visual analysis of immediacy of change in SSRD



Degree of Overlap (Effect Size)

In addition to visual analysis, nonoverlap of all pairs (NAP), is calculated as a non-parametric measure of effective size (Parker & Vannest, 2009). It compares each individual datum point in the baseline conditions to the subsequent treatment conditions to calculate the probability that a random score drawn in baseline would overlap a random score drawn in the treatment phase. In other words, NAP calculates the percentage of non-overlapping data between baseline and treatment phases. Effect sizes calculated via NAP scores are characterized as weak (0 – 0.65), moderate (0.66 – 0.92), or large (0.93 – 1.00) (Parker & Vannest, 2009).

Measuring the degree of overlap is important, given that real world data is often not as clean and clear. While the analysis of level, trend, variability, and immediacy of effect are important, obtaining an effect size provides further evidence of a strong or weak relationship between the independent and dependent variable.

The interpretation of this data provided insight into ways in which single subject design can be used to analyze the effects of cohort and program initiatives using both visual analysis and statistical analysis

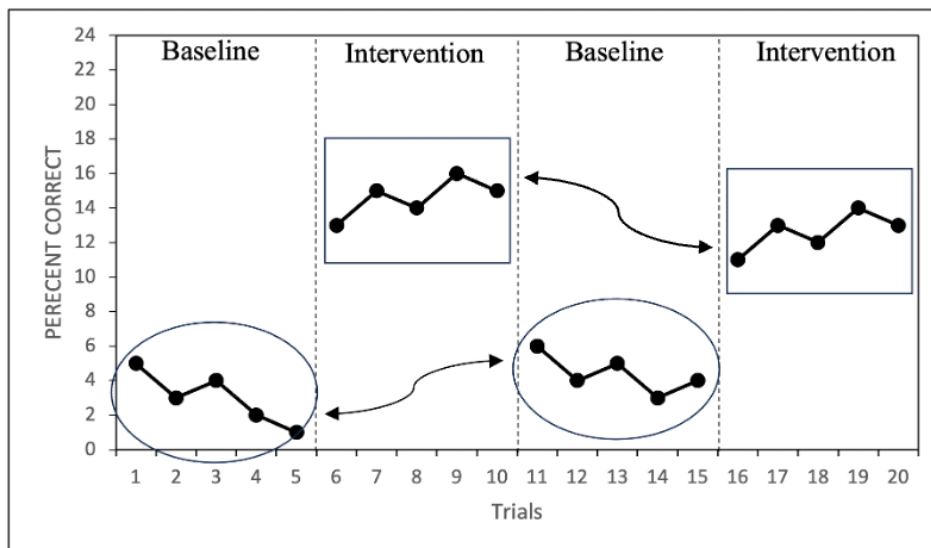
(NAP). Additionally, it has highlighted how 25 participants can be categorized into an N of 1 to evaluate program-level effectiveness and could also be categorized into an N of 4 (freshman, sophomore, junior, senior) to evaluate cohort-level effectiveness with no need for randomization. Results of this analysis additionally speak to the practicality of these analyses when considering the effectiveness of programs initiated in higher education.

Consistency of Data Patterns

Finally, consistency of data patterns involves looking at data from all phases within the same condition (e.g., all “A” conditions, all “B” conditions) and examining the extent to which there was consistency in these data patterns (Kratochwill, et al., 2010). This aspect of visual analysis, like in the hypothetical example below (Figure 6), provides insight into the ability to replicate the change in the dependent variable. In this example, when the independent variable was not present in baseline it was observed that the data set remained around the same level, trend and variability. In both instances when the independent variable was introduced it was observed that the data rose to similar levels with a similar trend and variability. Additionally, the replication of effect, both in baseline and in intervention, provides further evidence of a relationship between both the independent and dependent variable.

Figure 6.

Sample visual analysis of consistency of data patterns in SSRD



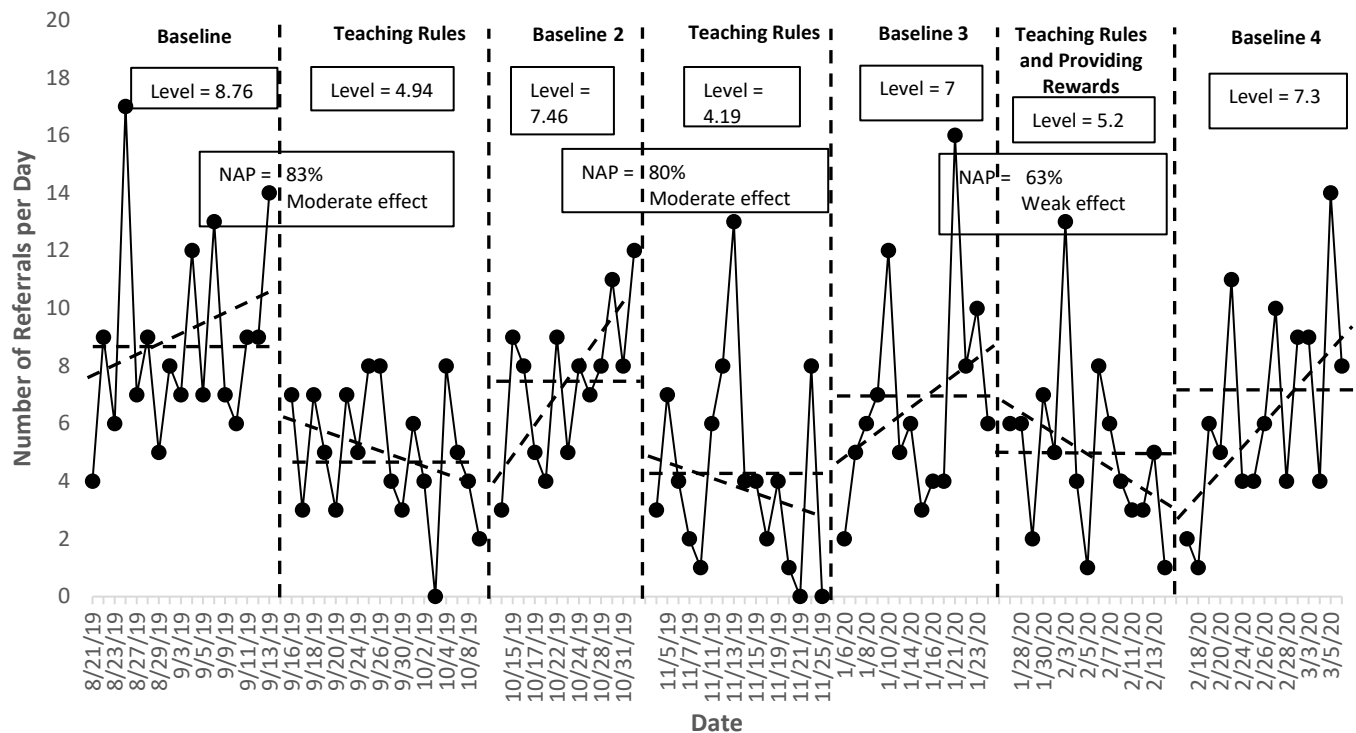
Results

Program-Level Results

As described, visual analysis of the results will include assessment of level, trend, variability, immediacy of change, degree of overlap, and consistency of data patterns (Horner et al., 2012 ; Kratochwill, et al., 2010) to evaluate effectiveness of the S.A.F.E. program in reducing the number of student referrals. A summary and visual analysis of program level results are reported in Table 5 and Figure 7.

Table 5.*Mean referrals, range, and effect size (NAP) for S.A.F.E. program*

Reversal Design Phase		Mean Referrals	Range	NAP	Effect Size Descriptor
A	Baseline (Round 1)	8.76	4-14	-	-
B	Teaching Rules (Round 1)	4.94	0-8	83	Moderate
A	Baseline (Round 2)	7.46	3-12	-	-
B	Teaching Rules (Round 2)	4.19	0-13	80	Moderate
A	Baseline (Round 3)	7.00	2-16	-	-
B+C	Teaching Rules & Providing Rewards	5.20	1-13	63	Weak
A	Baseline (Round 4)	7.30	1-14	-	-

Figure 7.*Number of total S.A.F.E. referrals (dependent variable) submitted per day***Level**

In baseline 1, the level (mean) of referrals given to students per day was 8.46, which decreased in the adjacent phase to 4.94 in teaching rules (round 1). In baseline 2, when staff discontinued teaching the rules to students, there was an increase in level to 7.46 referrals per day, which dropped again to a level of 4.19 in teaching rules (round 2). In baseline 3, staff discontinued teaching rules and the level increased once more to 7 referrals per day. Staff then began teaching rules and providing rewards for

following the rules and the level of referrals decreased to 5.2. In baseline 4, staff discontinued teaching rules one last time and the level of referrals increased again to 7.3 per day.

Trend

In baseline 1, when the rules were not taught there was a moderate increasing trend in the number of referrals received per day. Once S.A.F.E. rules were taught in teaching rules (round 1), there was a moderate decreasing trend over time. In baseline 2, there was an observed strong increasing trend in referrals received per day after staff discontinued teaching the rules. A moderate decreasing trend was observed once again in teaching rules (round 2). In baseline 3, the moderate positive trend in referrals per day returned. When the staff began teaching rules and providing rewards a strong decreasing trend was observed. A strong positive trend reoccurred in baseline 4 when the rules and rewards were discontinued.

Variability

A high level of variability was observed across all baseline and intervention conditions. Early on in this intervention it was determined that the high level of variability was a result of variable student schedules associated with college class schedules. Given the consistent variability it was decided that no adjustments would be made to the program.

Immediacy of Change

In analyzing the change in the last three data points in baseline 1 and the first three data points in teaching rules (round 1) there was an immediate decrease in the number of referrals with no overlap of data points. When analyzing the immediacy of change between teaching rules (round 1) and baseline 2 it was observed that the first data point in baseline 2 overlapped with the last three data points in teaching rules (round 1). An immediate change was observed between baseline 2 and teaching rules (round 2). In analyzing the first and last three data points between teaching rules (round 2) and baseline 3 it was observed that all three data points in baseline 3 overlapped with one data point in teaching rules (round 2). In analyzing the immediacy of change between baseline 3 and teaching rules plus rewards it was observed that two data points in the teaching rules plus rewards condition were equal to the final data point in baseline 3. Lastly, when analyzing the first and last three data points between teaching rules and baseline 4 it was observed that two data points in baseline 4 overlapped with the last three data points in teaching rules plus rewards.

Degree of Overlap (Effect Size)

When calculating NAP, each baseline is compared to the adjacent condition that follows it. The NAP score between baseline 1 and teaching rules (round 1) was 83%, indicating a moderate effect. The calculated NAP between baseline 2 and teaching rules (round 2) was 80% and was also considered a moderate effect. In the final score between baseline 3 and teaching rules plus rewards, NAP was calculated to be 64% and considered to be a weak effect.

Consistency of Data Patterns

Across all baseline conditions, visual analysis indicated that baseline conditions consistently showed an increase in level and trend, while all treatment phases, including the implementation of rules and rewards, resulted in a decrease in level and trend. Additionally, level, trend, and variability during all baselines and all treatment conditions were similar in value and presentation.

Discussion

Given that the intention of this intervention was to decrease the number of student referrals received per day, this SSRD revealed that S.A.F.E. was overall effective at the program level. Visual analysis showed a clear drop in the average number of referrals per day, and over time, when S.A.F.E. treatment conditions (i.e., teaching rules, providing rewards) were in place. When S.A.F.E. was removed, the number of referrals per day were higher on average and tended to increase over time. These results gave a strong indication that the implementation of S.A.F.E. had a functional relationship with the number of student referrals received per day.

One particularly noteworthy finding was that the average number of referrals received per day were higher, on average, when the program implemented “teaching rules and rewards” (treatment condition B + C) versus “teaching rules” alone (treatment condition B). This suggests that contrary to what one might expect, the combination of teaching rules and providing rewards was not more effective than just teaching the rules of S.A.F.E. alone. This conclusion was bolstered by NAP effect sizes, which indicated a moderate effect size for teaching rules (83% and 80% for rounds 1 and 2 of the teaching rules condition B) and a weak effect size for the combination of teaching rules and providing rewards (63% for treatment condition B + C). In addition to offering insights regarding effectiveness of S.A.F.E. program, we also highlight the ways in which this study supports the potential for SSRD to offer a rigorous, responsive, and accessible approach to assessment and evaluation in higher education.

Rigor

This paper highlights the potential to apply a rigorous, experimental design to the evaluation of a small number of students ($N=25$) without the need for randomization. While the current dataset includes a population of students that happen to have disabilities, almost any other program in higher education or student affairs could be substituted in its place.

While Mayhew et al. (2016) suggested that the measurement of the suggested programming above would be too difficult to obtain rigorous experimental control, as college settings create challenges with isolating participants and creating experimental and control groups, SSRD allows experimenters to obtain experimental control by using the participant(s) as their own control, essentially eliminating the need to randomize or isolate participants. Researchers and practitioners alike would only need to collect data on the dependent variable prior to the implementation of a program to begin the process of measuring change. Ultimately, experimental control is determined by the replication of the effect on the dependent variable, which can be accomplished using a multitude of different designs created to best match the research question.

This has potential implications for expanding robust assessment methods to minoritized populations of students (e.g., Native Americans), who are often classified as “other,” thereby negating their experiences and silencing their voices in higher education (Curley, 2019; Faircloth et al., 2015; Vaccaro et al., 2015). Thus, SSRD could speak to a growing movement toward critical quantitative approaches to research and assessment – all while advancing the field beyond reliance on purely correlational research designs (Wells & Stage, 2015).

Responsiveness

SSRD as a data collection method also provides practitioners with a useful tool for real-time data-driven decision-making. Ongoing use of visual analysis, in particular, guides practitioners in quickly responding to immediate student needs via data-driven allocation of resources (Horner et al., 2012). For example, although the data in this study was used primarily to evaluate program-level performance, it could also be leveraged to inform decisions about individual students, environments where referrals were most common, and types of referrals most frequently received. The consistent measurement of the dependent variable across time has the potential to inform a program director that one cohort may need different programming than another cohort or that more resources need to be allocated to one group versus another, which is exemplified in this study’s data set by the difference in referrals collected for the four separate cohorts in the program. For example, the consistent measurement of the dependent variable across time has the potential to inform a program director that one cohort may need different programming than another cohort or that more resources need to be allocated to one group versus another. Those analyzing the data might also determine that additional attention and resources are needed for senior students, while freshman students do not require as much support. This represents data-driven decision making that can be accomplished in real time. Traditional post hoc approaches to higher education research and assessment (Mayhew et al., 2016; Wells et al., 2015) lack such responsiveness.

Accessibility

Perhaps most importantly, this study’s results were achieved by practitioner-level professionals (university staff and students) in an unaltered unit of higher education. This study included limited professional staff with the majority of data collectors being part-time undergraduate students employed in the program. In a field where assessment, evaluation, and research competencies are often lacking among practitioners (e.g., Sriram & Oster, 2012; Sriram, 2014), SSRD is a promising option to guide data-driven decision-making among higher education professionals with a wide range of research skills and interests.

Implications for Future Use of SSRD in Higher Education

While higher education has not yet realized the full potential of SSRD, researchers from other disciplines have recognized its value and have provided evidence of its utility and effectiveness. This can specifically be observed at the student level, program level, and university level of operation.

Student-Level Learning and Support

At the student level SSRD in higher education would likely benefit most from increased capabilities in evaluating approaches to improving students’ performance in the classroom. For example, Messling

and Dermer (2010) used SSRD to show that 45 undergraduate students' attendance and preparation for class increased when showing up to class and showing proof of reading earned the right to use notes on their tests. Bicar et al. (2012) used SSRD to find that requiring 4 college athletes to text an academic counselor when they got to class increased their attendance and punctuality. Additionally, Farros et al. (2020) used SSRD to evaluate the effects of synchronous discussion within an asynchronous class on 3 graduate students quiz and participation performance, finding that there was no difference in grades, but a much higher level of participation during synchronous discussions. In another example, SSRD has shown that the use of guided notes can increase the participation and recall of information in undergraduate psychology students (Austin, et al., 2004).

Similar to these examples of student-level learning and support, SSRD holds promise as a useful tool in higher education where faculty and student affairs professionals are asked each semester to serve and teach students. This approach to evaluation would allow faculty to evaluate teaching approaches throughout a semester to find out what increases engagement or improves performance and for whom. Perhaps SSRD uncovers specific programming that impacts variables like attendance or persistence or graduation for students of a specific race, ethnicity, gender identify or sexual preference. It is also possible that this method of data collection uncovers that certain programming is best utilized only at specific times within the semester, such as just before midterms or finals.

Program-Level Evaluation and Decision-Making

While classroom performance is obviously important and spans a wide range of research questions, higher education is comprised largely of different programs and units, which SSRD also has the capability of evaluating. Foster (2010), for example, suggests SSRD could be used to measure the effectiveness of counseling with clients, providing multiple ways in which counselors and programs could evaluate change. Yet, while some examples highlighting the use of SSRD in higher education exist, the versatility of this research design holds the potential to be more broadly applied on college campuses by researchers and practitioners alike. For example, career centers would be interested in a study by Barker et al. (2019), which found that immediate feedback was more effective than post-session feedback in improving 6 undergraduate students' job interview skills. Smith and Ward (2006) even used SSRD to evaluate different types of feedback on the performance of 3 college football wide receivers, additionally looking to see if what was practiced by those players was used in games. Lastly, Wilk and Redmon (1998) and Tittelbach et al. (2007), as previously mentioned, used SSRD to evaluate goal setting and feedback on the performance of higher education admissions staff and student workers in a student advisors' office, respectively.

In the context of higher education, these approaches to evaluating program-level outcomes could translate to specific TRIO programs or veteran's affairs; for example, university administrators could examine the effectiveness of implemented programming (e.g., tutoring initiative, guest speaker series, etc.) by tracking their students' grades across a semester or their persistence to graduation across their college career. Fraternities and sororities could examine different approaches and their effectiveness at reducing alcohol consumption among its members by tracking the number of referrals made to the dean of students. Universities could examine specific programs effectiveness at addressing mental health by tracking their waitlists in student counseling or referrals to the dean of students.

University-Level Evaluation and Decision-Making

SSRD can also be used at the university-level for research questions related to broader initiatives. Reetz et al. (2014) evaluated the effects of posted signage and verbal reminders on students' bicycle and pedestrian lane adherence on a college campus footbridge, finding verbal reminders to be slightly better. Similarly, SSRD has been used to evaluate a program to reduce driving on campus for 12 college students on campus, ultimately finding a 20% reduction (Foxy & Hake, 1977). Additionally, Kazbour and Bailey (2010) used SSRD to evaluate the provision of incentives in a college bar to increase the percentage of people acting as or riding with a designated driver, finding that advertising free pizza and gas outside of the bar was most effective. Lastly, Clayton and Blaskewicz (2012) investigated the use of different types of signage to decrease urinal splatter in men's bathrooms on a college campus and reduce the number of resources needed for nightly cleaning. Ultimately the use of a sign and target installed in the urinal led to significant reductions.

Within the context of higher education, practitioners are tasked with a multitude of services related to care, entertainment, growth and development, and wellness, just to name a few. In real time universities could evaluate mental wellness programming and initiatives to see what decreases student counseling waiting lists or more importantly, mental health crises in specific student populations. Such applications of SSRD could then be used inform university-level decision-making related to students' needs and the effectiveness of university services in meeting those needs.

Conclusion

This study provides an illustrative example of SSRD in the context of higher education assessment, evaluation, and research practices. This example highlights the promise SSRD holds in offering researchers and practitioners an additional innovative assessment design that addresses many of the challenges currently impeding effective AER practices in higher education. As demonstrated in this study, SSRD can serve as a useful complement to existing assessment and evaluation practices. Ultimately, SSRD can further develop a robust evidence base to inform student-level learning and support, as well as program-level and university-level evaluation and decision-making, across colleges and universities.

In particular, this study highlighted the ways in which SSRD can serve as a rigorous, responsive, and accessible approach to data-driven decision-making in higher education. As highlighted throughout this example, SSRD provides researchers and practitioners with robust, experimental designs with which they can answer questions regarding who interventions are effective for and under what circumstances those interventions are effective (rigor). Moreover, by providing real-time analysis of interventions in practice, SSRD can provide an approach to AER that is sensitive to change, thereby allowing practitioners and researchers to quickly to changing student needs (responsive). Finally, the user-friendly design of SSRD offers administrators and practitioners an effective method for making informed, data-based decisions in real time, regardless of their research skills and training (accessible).

References

- Adelman, C. (1999). *Answers in the toolbox: Academic intensity, attendance patterns, and bachelor's degree attainment*. U.S. Department of Education. Retrieved on March 9, 2022, from <http://www.ed.gov/pubs/Toolbox/index.html>
- American College Personnel Association & NASPA-Student Affairs Administrators in Higher Education. (2010). *Professional competency areas for student affairs practitioners*.
- Austin, J. L., Lee, M., & Carr, J. P. (2004). The Effects of Guided Notes on Undergraduate Students' Recording of Lecture Content. *Journal of Instructional Psychology*, 31(4).
- Barker, L. K., Moore, J. W., Olmi, D. J., & Rowsey, K. (2019). A comparison of immediate and post-session feedback with behavioral skills training to improve interview skills in college students. *Journal of organizational behavior management*, 39(3-4), 145-163. <https://doi.org/10.1080/01608061.2019.1632240>
- Belkin, D. (2024, January). Why Americans have lost faith in the value of college. *The Wall Street Journal*. <https://www.wsj.com/us-news/education/why-americans-have-lost-faith-in-the-value-of-college-b6b635f2>
- Bicard, D. F., Lott, V., Mills, J., Bicard, S., & Baylot-Casey, L. (2012). Effects of text messaged self-monitoring on class attendance and punctuality of at-risk college student athletes. *Journal of Applied Behavior Analysis*, 45(1), 205-210. <http://dx.doi.org/10.1901/jaba.2012.45-205>
- Behavior Analyst Certification Board. (2024, April 1). *Organizational Behavior Management*. <https://www.bacb.com/about-behavior-analysis/organizational-behavior-management/>
- Blimling, G. S. (2001). Uniting scholarship and communities of practice in student affairs. *Journal of College Student Development*, 42(4), 381-396.
- Braxton, J. M., & Hirschy, A. S. (2005). Theoretical developments in the study of college student departure. *College Student Retention: Formula for Student Success*, 3, 61-87.
- Carpenter, S. (2001). Student affairs scholarship (re?)considered: Toward a scholarship of practice. *Journal of College Student Development*, 42(4), 301-318.
- Clayton, M. C., & Blaskewicz, J. (2012). The use of visual prompts to increase the cleanliness of restrooms on a college campus. *Journal of Organizational Behavior Management*, 32(4), 329-337. <https://doi.org/10.1080/01608061.2012.729393>
- Cooper, T., & Terrell, T. (2013). *What are institutions spending on assessment? Is it worth the cost?* (Occasional Paper 18). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA). <http://learningoutcomesassessment.org/documents/What%20are%20institutions%20spending%20on%20assessment%20Final.pdf>
- Cox, B.E., Reason, R.D., Tobolowsky, B.F., Brower, R.L., Patterson, S., Luczyk, S. & Roberts, K. (2017) Lip service or actionable insights? Linking student experiences to institutional assessment and data-driven decision making in higher education, *The Journal of Higher Education*, 88(6), 835-862, <https://doi.org/10.1080/00221546.2016.1272320>
- Curley, K. M. (2019). Methods to re-center the “other”: When discarding outliers means discarding already marginalized stories. *Journal of Critical Thought and Praxis*, 8(2), 104-114. <https://doi.org/10.31274/jctp.8795>
- Erath, T. G., Pellegrino, A. J., DiGennaro Reed, F. D., Ruby, S. A., Blackman, A. L., Novak, M. D. (2021). Experimental research methodologies in Organizational Behavior Management, *Journal of*

- Organizational Behavior Management*, 41(2), 150-181, <https://doi.org/10.1080/01608061.2020.1869137>
- Fairchild, L.A., Powell, M. B., Gadke, D. L., Spencer, J. C., & Stratton, K. K. (2020). Increasing social engagement among college students with autism using peer mentorship: A pilot study. *Advances in Autism*, 6(2), 83-93.
- Faircloth, S. C., Alcantar, C. M., & Stage, F. K. (2015). Use of large-scale data sets to study educational pathways of American Indian and Alaska Native students. *New Directions for Institutional Research*, No. 163, 5-24.
- Farros, J. N., Shawler, L. A., Gatzunis, K. S., & Weiss, M. J. (2020). The effect of synchronous discussion sessions in an asynchronous course. *Journal of behavioral education*, 1-13. <https://doi.org/10.1007/s10864-020-09421-2>
- Foster, L. H. (2010). A best kept secret: Single-subject research design in counseling. *Counseling Outcome Research and Evaluation*, 1(2), 30-39. <https://doi.org/10.1177/2150137810387130>
- Foxx, R. M., & Hake, D. F. (1977). Gasoline conservation: A procedure for measuring and reducing the driving of college students. *Journal of Applied Behavior Analysis*, 10(1), 61-74.
- Fuller, M., Henderson, S., & Bustamante, R. (2015). Assessment leaders' perspectives of institutional cultures of assessment: A Delphi study. *Assessment & Evaluation in Higher Education*, 40(3), 331-351. <http://dx.doi.org/10.1080/02602938.2014.917369>
- Hersh, R. H., & Keeling, R. P. (2013, February). *Changing institutional culture to promote assessment of higher learning*. (Occasional Paper No. 17). University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA). <https://www.learningoutcomesassessment.org/wpcontent/uploads/2019/02/OccasionalPaper17.pdf>
- Higher Education Opportunity Act-2008. (2010, June 28). *Laws and guidance - Higher education*. U.S. Department of Education. Retrieved June 21, 2022, from <https://www2.ed.gov/policy/highered/leg/hea08/index.html>
- Hoffman, J. L., & Bresciani, M. J. (2010). Assessment work: Examining the prevalence and nature of learning assessment competencies and skills in student affairs job postings. *Journal of Student Affairs Research and Practice*, 47(4), 495-512.
- Horner, R. H., Swaminathan, H., Sugai, G., & Smolkowski, K. (2012). Systematic analysis and use of single-case research. *Education and Treatment of Children*, 35(2), 269-290.
- Hong, R. C., & Moloney, K. (2020, October). There is no return to normal: Harnessing chaos to create our new assessment future. (Occasional Paper No. 49). University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Institute of Education Sciences. (n.d.). *What Works Clearinghouse: Frequently asked questions*. <https://ies.ed.gov/ncee/wwc/FAQ>.
- Jaschik, S., & Lederman, D. (2020). *College and university chief academic officers*. Inside Higher Ed and Gallup.
- Kazbour, R. R., & Bailey, J. S. (2010). An analysis of a contingency program on designated drivers at a college bar. *Journal of applied behavior analysis*, 43(2), 273-277. <http://dx.doi.org/10.1901/jaba.2010.43-273>
- Kratochwill, T., Hitchcock, J., Horner, R., Levin, J., Odom, S., Rindskopf, D., & Shadish, W. (2010). *Single case designs technical documentation*. <https://files.eric.ed.gov/fulltext/ED510743.pdf>

- Marsh, G. (2014). Institutional characteristics and student retention in public 4-year colleges and universities. *Journal of College Student Retention: Research, Theory and Practice*, 16(1), 127–151. <https://doi.org/10.2190/CS.16.1.g>
- Mayhew, M., Rockenbach, A., Bowman, N. Seifert, T., Wolniak, G., Pascarella, E., & Terenzini, P. (2016). *How college affects students: 21st century evidence that higher education works*. Jossey-Bass.
- Messling III, P. A., & Dermer, M. L. (2010). Increasing students' attendance at lecture and preparation for lecture by allowing students to use their notes during tests. *The Behavior Analyst Today*, 10(3-4), 381.
- No Child Left Behind Act of 2001, Pub. L. No. 107–110, 115 Stat. 1425 (2002).
- Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior therapy*, 40(4), 357-367. <https://doi.org/10.1016/j.beth.2008.10.006>
- Reetz, N. K., Loukus, A. K., Taylor, K. M., & Dixon, M. R. (2014). Investigating the effects of mediated and passive prompts on pedestrian and bicycle lane adherence on a college campus footbridge. *Journal of Organizational Behavior Management*, 34(2), 144-155. <https://doi.org/10.1080/01608061.2014.914008>
- Roane, H. S., Fisher, W. W., & Carr, J. E. (2016). Applied behavior analysis as treatment for autism spectrum disorder. *The Journal of Pediatrics*, 175, 27-32. <http://dx.doi.org/10.1016/j.jpeds.2016.04.023>
- Ross, J. (2009). Visual perception 1950–2000. *Inside psychology: A science over 50 Years*, 243–252.
- Seidman, A. (2005). Minority student retention: Resources for practitioners. *New Directions for Institutional Research*, 2005(125), 7-24.
- Smith P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin and Review*, 25, 2083-2101. <https://doi.org/10.3758/s13423-018-1451-8>
- Smith, S. L., & Ward, P. (2006). Behavioral interventions to improve performance in collegiate football. *Journal of applied behavior analysis*, 39(3), 385-391. <http://dx.doi.org/10.1901/jaba.2006.5-06>
- Sriram, R. (2014). The development, validity, and reliability of a psychometric instrument measurement competencies in student affairs. *Journal of Student Affairs Research and Practice*, 4, 349-363.
- Sriram, R., & Oster, M. (2012). Reclaiming the “scholar” in scholar-practitioner. *Journal of Student Affairs Research and Practice*, 49(4), 377-396. <https://doi.org/10.1515/jsarp-2012-6432>
- Tittelbach, D., DeAngelis, M., Sturmey, P., & Alvero, A. M. (2007). The effects of task clarification, feedback, and goal setting on student advisors' office behaviors and customer service. *Journal of Organizational Behavior Management*, 27(3), 27-40. https://doi.org/10.1300/J075v27n03_03
- Vaccaro, A., Kimball, E. W., Wells, R. S., & Ostiguy, B. J. (2015). Researching students with disabilities: The importance of critical perspectives. *New Directions for Institutional Research*, No. 163, 25-42.
- Wells, R. S., & Stage, F. K. (2015). Past, present, and future of critical quantitative research in higher education. *New Directions for Institutional Research*, No. 163, 103-112.

- Wilk, L. A., & Redmon, W. K. (1998). The effects of feedback and goal setting on the productivity and satisfaction of university admissions staff. *Journal of Organizational Behavior Management*, 18(1), 45-68.
- Young, R. B. (2001). A perspective on the values of student affairs and scholarship. *Journal of College Development*, 42(4), 319-337.

About the Authors

Jordan Spencer, Director of ACCESS and Specialized College Services, Mississippi State University, js5152@msstate.edu

Christa Winkler, Assistant Professor, Mississippi State University, cw2979@msstate.edu

Appendix

Level	Trend	Variability	Immediacy of Effect	Consistency of Data Patterns
2	5	6	2	5
3	3	4	3	3
2	4	5	2	4
3	2	3	3	2
2	1	4	2	1
14	13	10	14	13
15	15	20	15	15
14	14	9	14	14
15	16	22	15	16
14	15	13	14	15
4	6	7	4	6
5	4	5	5	4
4	5	6	4	5
5	3	4	5	3
4	4	5	4	4
17	11	9	17	11
18	13	21	18	13
17	12	11	17	12
18	14	19	18	14
17	13	10	17	13