Single Case in the Social Sciences 1(1)
DOI: 10.15763/issn.1936-9298.2024.1.1.15-35
http://singlecasejournal.com/

Comparing the Effectiveness of Restricted-Operant and Free-Operant Teaching Arrangements on Measures of Acquisition and Fluency Outcomes

Andrew J. Bulla[1], Jennifer L. Wertalik[1], and Leah Yakabovits[1]

Abstract
Discrete-trial teaching (DTT), a restricted-operant teaching arrangement, and frequency-building instruction (FBI), a free-operant teaching arrangement, represent two instructional strategies derived from operant conditioning. Researchers and practicing behavior analysts have used both to establish and firm up novel stimulus-behavior relations. Despite the effectiveness of both procedures, few studies have compared the two techniques and assessed the effects on the emergence of fluent responding. The current study extends the research to typically developing college students to directly compare DTT and FBI. We taught participants the numerals 0-10 in unknown foreign languages (i.e., Mandarin, Arabic, and Hindi) using both procedures. Under both conditions, we held constant the number of practice trials and frequency of reinforcement. Results found quicker acquisition under restricted-operant conditions for all participants, though none of the participants met the desired frequency aim prior to the end of the study. We discuss the results of the study in the context of planning for learning across three stages of learning.

Keywords
free operant, restricted operant, standard celeration chart, adapted alternating treatments design

[1]Department of Psychology, Georgia Southern University – Armstrong Campus, Savannah, GA, USA

Corresponding Author:
Andrew J. Bulla, Department of Psychology, Georgia Southern University – Armstrong Campus, Savannah, GA, United States of America, 31419. E-mail: abulla@georgiasouthern.edu

Researchers have suggested that teaching paradigms often exist on a continuum from restricted operant (RO) to free operant (FO; Bulla, 2023; Evans et al., 2021). Restricted operant paradigms place an external restriction on the frequency of responses emitted by an organism, most often in the duration of the inter-trial-interval (ITI). For example, the instructor provides an S^D, "touch your nose," the learner touches their nose, the instructor provides a reinforcer, and then an ITI of 15 s occurs prior to the next S^D (Ghezzi, 2007). This typically occurs for a predetermined number of trials and yields a statistic about the accuracy of the performance. Free operant paradigms seek to lift any restriction on the natural frequency of a behavior, allowing the organism to freely respond numerous times in succession, seeking to eliminate imposed ITIs. For example, given the same skill of listener responding, the instructor presents the S^D, "touch your nose," the learner touches their nose, and immediately after the learner engages in the behavior, the instructor delivers the next S^D, thus reducing the ITI. This occurs for a predetermined length of time and yields a frequency of corrects and errors per unit of time. Under this paradigm, one places the focus on increasing the frequencies of the behavior.

Researchers and behavior analysts have developed a wide range of instructional technologies based on these paradigms. Discrete-trial teaching (DTT) represents one of the most used teaching procedures derived from restricted-operant paradigms (Ghezzi, 2007), and frequency-building instruction (FBI) representing one of the most common derived from free-operant paradigms (Gist & Bulla, 2022). Both RO and FO teaching arrangements provide learners with a research-based, effective instructional arrangement for building accurate performance (Ghezzi, 2007; Gist & Bulla, 2022). A free-operant teaching paradigm aims to remove ceilings on learner performance by providing instructional arrangements that support the freedom to perform at rates and topographies restricted only by the laws of nature (Lindsley, 1996). This paradigm also allows for direct and continuous measurement of response rates— the data from which the analysts then use as the primary dependent variable. Conversely, in a restricted operant arrangement, teachers present materials prompting one response at a time with an ITI separating response opportunities (Ghezzi, 2007), allowing the learner more time to contact stimulus conditions prior to making a response. The resulting data analysis typically occurs with a derived measurement such as percent correct as the primary datum.

While the robust literature exists on the effectiveness of DTT procedures (Ghezzi, 2007), such interest focusing on the state of effectiveness of methodologies derived from FO paradigms have recently emerged (Gist & Bulla, 2022). Doughty, Chase, and O'Shields (2004) conducted a literature review on the effectiveness of frequency-building procedures. Researchers assessed the level of supporting research on the use of frequency building across a variety of skills (e.g., spelling). Results identified two variables that studies did not consistently control for, thus weakening the quality of research on frequency-building procedures. First, a lack of appropriate controls for reinforcement across conditions existed, with frequency-building procedures often resulting in more frequent reinforcement. Second, a lack of appropriate controls for the overall number of practice trials across conditions also existed, with frequency building often resulting in more practice opportunities compared to DTT.

Doughty et al (2004) also discussed the need for more objective measures for the associated outcomes of fluency. They argued that a consistent definition for the outcomes of fluency did not exists in the literature. Precision teachers often look for outcomes related to maintenance, endurance, stability, application, and adduction (MESAA; Johnson et al., 2020). One observes *maintenance* when the same or better frequency of performance occurs after a period of no practice. *Endurance* represents the same performance when the length of observation increases (e.g., same performance during a 10-s timing and a 60-s timing). *Stability* represents the same frequency of performance in the face of distracting stimuli or contexts. *Application* occurs when the learner uses the same taught skill under novel contexts that resemble the teaching contexts. Lastly, *adduction* occurs when the learner engages in an untaught behavior, most often a combination of established behaviors, in a novel context. Previous research suggest that frequencies must maintain at the same level across these areas for the behavior to be considered "fluent" (Binder, 1993). Though precision teachers test for these outcomes, the published literature does not always include each of the components in MESAA, making it difficult to make comparisons of "fluent" behavior. For example, Aravamudhan & Awasthi (2021) included measures of maintenance, endurance, stability, and application after using FBI to increase speech sounds in an autistic female but did not include measures of adduction.

Regardless of the instructional technique, both paradigms seek to establish skills that maintain across time and occur at a level that allows the environment to naturally recruit these established behaviors under the naturally occurring stimulus conditions. Researchers have begun to investigate the performance standards necessary for such mastery to occur. Richling et al. (2019) conducted a practitioner survey and found that most ABA clinicians use 80% accuracy across three sessions as mastery criteria. Richling et al. (2019) also found that only those skills taught with 100 % accuracy as mastery criteria produced retention between 80% to 90 % whereas skills built to lower criteria deteriorated substantially. However, other studies associate the ability to respond accurately at high rates with fluency outcomes such as retention, endurance, stability, application, and adduction (Fabrizio & Moors, 2003; McTiernan et al., 2016; Stocker et al., 2019), driving fluency-based instructors to aim for high rates of responding as mastery criteria.

Researchers have also begun to include the appropriate controls for the abovementioned variables in current investigations. Noppraprun and Holloway (2014) found mixed results in their investigation of DTT and FBI to teach letter sounds to individuals with Autism Spectrum Disorder, while controlling for practice opportunities and the rate of reinforcement, and comparing instructional duration, sessions to accuracy, and rate of response in fluency checks. The DTI condition produced accuracy in fewer sessions for two participants, whereas rate building in the FT condition produced accuracy in fewer sessions for two additional participants as well as a more efficient instructional arrangement (i.e., shorter overall duration) for all participants, and better fluency outcomes for all participants in two of the fluency checks. This provides evidence that while students may learn novel skills with both DTT and FBI, better long-term effects may occur with frequency building.

The present study investigated the effects of free-operant (FO) and restricted-operant (RO) teaching arrangements on the acquisition of novel discriminations with adult learners in a translational setting. We compared the rate of acquisition with foreign language numerals (Hindi, Arabic, and Mandarin), and fluency outcomes attained in RO, FO, and control (no teaching) conditions with college students. Fluency outcomes included measuring endurance, stability, application, and adduction based on the procedures outlined by Fabrizio and Moors (2003) and Johnson and Street (2012). The emergence of the global COVID-19 pandemic prevented conducting maintenance checks or social validity measures.

Method

Participants and Setting

Table 1 presents the demographic data for the six college students that participated in this study. Participants included within the study responded to either a recruitment email or through signed up to participate via SONA, the online resource for research recruitment within the university. All participants received extra credit through an online subject pool for undergraduate students. The researchers provided the participants' instructors with the total amount of hours the individual participated in the study, and the instructor used that information to distribute extra credit. Researchers excluded one participant, Claire, from the study. During the control condition (i.e., no instruction), researchers detected inexplicable gains in Claire's performance indicative of a history threat—namely, practice outside of the research setting. Unmonitored practice, confirmed by the participant, negated the effects of the relevant teaching interventions in the study; therefore, a total of five participants remained for analysis.

Sessions included one participant and one researcher. All sessions took place in the same 9' X 12' room with a table and two chairs. An additional room attached to the lab room provided researchers the opportunity to conduct treatment integrity through a one-way window without disrupting the session. Each participant engaged in a baseline session to ensure they did not have the discrimination skills already, and then approximately two sessions per week across 5 weeks for a total of 10 – 11 additional sessions, followed by the MESAA checks during the final session.

Table 1. Participant Demographics.

Participant ID	University Class	Age	Gender	Race	Combined Family Income
Jill	Sophomore	20	Female	Black/African American	$75,000-$100,000
Leon	Junior	26	Male	White/Caucasian	$25,000-$50,000
Chris	Sophomore	27	Male	Black/African American	$25,000-$50,000
Claire	Sophomore	20	Female	Latina-Hispanic	$50,000-$75,000
Ada	Sophomore	20	Female	White/Caucasian	< $25,000
Sheva	Junior	19	Female	White/Caucasian	$75,000-$100,000

Materials

Stimuli used in the RO condition included three sets of 300 white colored, 3-inch x 5-inch index cards with a character numeral (1–10) centered on the front in 48- point, black font, in one of three languages: Mandarin, Arabic, or Hindi. The stimuli used in the FO condition included 24 worksheets— eight for each language— with randomized numerals (1–10) listed in 20 rows and 10 columns, totaling 200 digits printed on standard (8.5-inch x 11-inch) white computer paper in black font. Researchers reserved four worksheets for each language to serve as novel stimuli for post-practice data collection probes.

Researchers used a standard timer for timings and marked correct answers with an audible clicking device. Datasheets contained columns indicating the trial number (25 trials per column, four columns) followed by a column with a (+) symbol and a (-) for selecting a correct (+) or incorrect (-) response on each trial. Data collectors used this same format of the data sheet for all language conditions during baseline, both intervention phases, and during MESAA check probes.

Dependent Variable

Researchers measured the frequency of correct and incorrect digits read aloud from a novel worksheet during a 30-s probe in each language. The probe worksheet included more stimuli than possible to read in 30 s to avoid imposing a ceiling on the learners' response rates. A correct response occurred when the participant said the correct numeral in English. An error occurred if an individual said, "pass," or an incorrect numeral. The researcher did not provide any feedback on measures of the dependent variable.

Procedures

During each condition, the participant sat in a chair at a table with the condition-specific teaching materials (i.e., RO cards or FO practice sheet) in front of them. The researcher sat in a chair across the table from the participant with a timer and data collection materials. At the conclusion of each session, researchers graphed data on a standard celeration chart for visual analysis and decision-making.

Baseline. Data collection in the baseline condition occurred for one session per participant to evaluate evidence of any previous experience with the foreign language numerals. During this condition, the participant sat at a table across from the researcher with the probe worksheet in front of them. Researchers instructed participants to say the numbers from the probe sheet as quickly as they could and to either guess or say, "pass" for unknown numbers. One timing occurred for each language and lasted for 30 s. A similar procedure occurred for the adduction check baseline, however, the probe sheet included random multi-digit numbers (11– 20). The application check baseline required participants to write the solutions of random single-digit math facts on a worksheet during a 30 s timing. Each probe sheet included math facts in one language and timings occurred for all three languages separately. The participants wrote the answers in Hindi-Arabic (i.e., the standard U.S.

numeral system). At the conclusion of the timings, the researcher thanked the participant and the baseline session concluded.

Control. In the control condition, participants engaged in the same timing procedure as during the baseline session with single digits in the language assigned to this condition. This condition served as the "no teaching" condition to control for the effect of improvement based on repeated practice with the probe procedure. Therefore, participants received no feedback on their performance or reinforcement of responses, and the condition included a single probe in the given language.

Free Operant

Phase 1. Researchers instructed participants to say the numbers from the worksheet as quickly as they could within a 30-s timing. The researcher marked correct responses by pressing a clicker and recorded the occurrence as a correct response on the data sheet. If the participant said the wrong number or said, "pass," the researcher gave an immediate echoic model of the correct number, paused for the participant to echo the correct response, and recorded the response as incorrect. When the participant echoed correctly within 3– 5 s, the researcher then pressed the clicker, and the participant continued to the next digit. This process continued for the duration of the 30-s timing. The researcher repeated the timing procedure for three, 30-s practice timings.

After the three practice timings, the researcher conducted the data collection probe to assess participants performance on the DV. Participants read from a novel probe sheet, data collection on correct and incorrect responses occurred, and researchers provided no error correction or reinforcement of correct responses.

Phase 2. When participants reached at least 90% accuracy in the FO and RO conditions during the same session, phase 2 commenced for both conditions. During phase 2 of the FO condition, three practice timings occurred as in phase 1, but no feedback occurred during the timings. Instead, at the end of each 30 s timing, the researcher stated how many correct and incorrect responses occurred and provided error correction by pointing to the incorrectly stated digit on the worksheet and saying the correct digit. Participants echoed the correct response before moving on to the next correction. If no errors occurred, the researcher stated that no errors occurred and congratulated the participant. This procedure occurred for three, 30-s practice timings before conducting the data collection probe using the same procedure as the probe in phase 1.

Restricted Operant Teaching

Phase 1. To determine the number of trials to conduct during the RO condition, researchers yoked the total number of responses that occurred in the FO practice timings for that session. Yoking response opportunities controlled for the potential impact of unequal practice opportunities that could occur from the different teaching arrangements. For example, if the participant said a total of 35 numerals (correct or incorrect) summed across all three FO practice sessions, the researcher presented a total of 35 stimulus cards.

The researcher placed the cards in a pile, then held up one card at a time in front of the participant, and delivered a verbal prompt (i.e., "What number?"). If the participant said the correct number, the researcher marked the response by pressing the clicker and recorded the responses as correct. If the participant said the incorrect number, said, "pass," or did not respond within three to five seconds, the researcher implemented the same error correction procedure as in the FO condition and marked the response as incorrect on the data sheet. Between all trials the researcher imposed a 3–5 s inter-trial interval (ITI) from the participant's response to the presentation of the next card. After the researcher presented all of the cards in RO training, the researcher conducted the data-collection probe exactly as the control condition probe with the language assigned to the RO condition.

Phase 2. As previously stated, when the participant achieved at least 90% accuracy in the RO and FO conditions during the same session, phase 2 began. During phase 2 of the RO condition, the participants completed three practice opportunities in a similar manner to phase 1 except the researchers did not deliver any feedback until the end of each practice opportunity. The researchers again yoked the number of stimulus presentations to match the number of practice opportunities presented in each of the FO timings. For example, if the participant said 60 digits for the first timing in the FO condition, the researcher presented 60 cards during the first block of practice in the RO training. If the participant said 50 digits during the second timing, the researcher presented 50 cards during the second block, etc. After participants completed the cards from one practice opportunity, the researcher informed the participant of the number of correct and incorrect responses. Error correction then occurred with the same echoic procedure as during phase 1, but researchers presented all errors consecutively with a 3–5 s ITI. If no errors occurred, the researcher stated that they made no errors and congratulated the participant. The same procedure occurred for all three RO practices. Researchers then conducted the data collection probe of the DV with the same procedure as in the FO condition.

MESAA Checks

Researchers planned to measure endurance, stability, application, and adduction according to MESAA check procedures described by Fabrizio and Moors (2003) when the participants reached a frequency aim of 120 correct numerals per minute. Unfortunately, the COVID-19 pandemic interrupted the study, making the participants inaccessible for continued sessions or retention checks which require a month with no practice. Therefore, MESAA checks occurred before participants met the frequency aim. MESAA checks occurred in each condition using a cold-probe procedure. Unlike phases 1 and 2, the cold-probe procedure included no practice opportunities, data collection occurred on the first timing of each check, and participants received no feedback.

Endurance. In the present study, data collection probes occurred during 30-s timings. Therefore, to probe for endurance researchers used the same procedures described previously except they increased the timing period to 90 s.

Stability. To test for stability of the skill in this study, participants engaged in 30-s timing while listening on headphones to a recording of the first author saying a series of randomized numbers in English. The response requirement remained the same as the data collection probes, with participants seeing digits in the condition-dependent language and saying the digit in English.

Application. For the application check, participants completed a 30-s timing with single-digit addition and subtraction problems in each of the condition-specific languages (i.e., Hindi, Arabic, and Mandarin) according to the same procedures as described in the baseline condition.

Adduction. Adduction represents a behavioral phenomenon in which novel, untrained instances of behavior emerge without explicit instruction (Johnson et al., 2020). To assess for adduction, researchers conducted a 30-s probe of untaught, multi-digit numerals (11 –20) in each language condition. Multi-digit identification in all three languages followed similar rules to standard English numbers. For example, the digit /13/ in all three languages had the numeral /1/ followed by the numeral /3/.

Experimental Design and Analysis

The present study employed an adapted alternating treatment design (AATD) to assess the effects of RO and FO on the accuracy, frequency, stability, application, and adduction of the dependent variable. Researchers commonly use an AATD when comparing the effectiveness and/or efficiency of two or more interventions for different but equally difficult target sets (Cariveau & Fetzer, 2022). When using an AATD, the researcher(s) assigns target sets to each condition and then introduces the conditions by rapidly alternating their presentation. In the present study, the researchers first conducted a logistical analysis to ensure instructional items consisted of equivalent target sets (Wolery et al., 2014). The researchers selected the three languages (i.e., Mandarin, Arabic, and Hindi) because each of these languages uses characters/symbols to represent numerals that differ from the numerical system used in the United States (e.g.,٢ = 2 in Arabic). The researchers independently analyzed each of the characters within each language to ensure that the symbols did not resemble traditional English numerals, all required rote memorization or simple stimulus control, and that teen numbers (e.g., 13) all consisted of a symbol used for “10” plus one additional symbol. The researchers also independently analyzed visual properties of the symbols within each language and deemed symbols within each language to be of equal size, shape, and complexity.

After equating target sets (i.e., character sets of numerals 1 – 10 in Mandarin, Arabic, and Hindi), the researchers randomly assigned each target set to three different conditions (i.e., FO, RO, and control) across participants using the randomization function within an Excel spreadsheet. The study consisted of an initial baseline condition to measure the participants performance prior to introduction to the two treatment conditions. A third target set, a control set, appeared in the baseline condition as well and received measurement throughout the experiment to assess threats to internal validity (Wolery et al., 2014, 2018). Furthermore, counterbalancing the target sets across conditions and participants provides an additional means of controlling for difficulty when using and AATD (Cariveau et al., 2022).

Each session consisted of all three conditions (i.e., RO, FO, Control) for each participant. Researchers then compared the effects of each condition to the control condition to observe differences in responding. To control for the number of practice opportunities across conditions, the number of trials obtained during each FO condition determined the number of trials presented during the RO condition in each session. Therefore, the FO condition occurred before the RO condition. Given the control condition did not vary in response opportunities during different sessions, participants could experience the control condition either before or after the FO condition. Researchers randomized the order of control and FO conditions prior to each session in a way that outcome produced a rank ordering of the three conditions. The researchers used the daily per minute standard celeration chart to evaluate the change in level, bounce, celeration, and accuracy improvement measures (Pennypacker et al., 2003).

Treatment Integrity

Before conducting each session, the researchers reviewed a treatment integrity checklist of procedural components. During random sessions, the primary investigator conducted an integrity check to ensure the researchers accurately implemented the intervention procedures. Observations for the integrity check took place behind a one-way mirror without the implementer's knowledge. Integrity checks occurred for 39% of the sessions, which included all conditions in both phases. Correct implementation occurred for an average of 99% of the steps, with a range of 93% – 100%. Researchers calculated total treatment integrity by dividing the total number of correctly implemented procedural components by the total number of components for each observed session.

Interobserver Agreement

Researchers assessed interobserver agreement (IOA) during 43% of total sessions across all participants during the baseline and intervention phases. Additionally, researchers collected IOA data across two participants sessions during the MESAA checks. During the assessment, a second researcher observed the participant's performance and collected data. Researchers then calculated trial-by-trial IOA by dividing the number of agreements by the total number of trials. A trial represented a response made by the participant to a relevant S^D (either on the sheet or delivered on cards). Baseline IOA included each condition for four out of five participants and resulted in average agreement of 98% (Range 87.5– 100). IOA across data collection probes in intervention averaged 98% (Range 95.8 – 100). MESAA check IOA included each MESAA check (i.e., endurance, stability, application, and adduction) during each condition for two out of five participants (40% of total MESAA checks) and achieved an average of 91.9% total agreement (Range 63.6 – 100). 63.6% occurred during the endurance probe of Ada's performance. The total agreement for Ada's MESAA checks averaged 85%, and the total agreement during Chris's MESAA checks averaged 98.75%.

Results

Standard Celeration Charts

Figures 1 –5 provide a visual display of the frequency, celeration, level, and bounce of correct and incorrect responses per minute for each participant in each condition on standard celeration chart (SCC) segments. Given charting conventions using the SCC (i.e., dots for acceleration data, x's for deceleration data), putting the data for all three conditions on one graph would not allow for proper visual analysis. Therefore, the researchers included three separate SCC segments, one for each condition, that appear next to each other and allow for visual analysis to take place by comparing data from left to right for each participant. When using an AATD, demonstration of experimental control occurs when differentiating in responding occurs between the different treatment conditions. By displaying the SCC segments in separate tiers, once can see the effects each condition had on behavior and this permits comparisons of the effects of the independent variables on the dependent variable.

Frequency describes the rate of responses occurring in a given time (Pennypacker et al., 2003). Celeration describes the growth or decay of a performance in a given celeration period (Pennypacker et al., 2003); This study used a one-week celeration period. Celeration value descriptions include a direction (i.e., X describes growth, whereas / describes decay) and an exponential magnitude of growth (e.g., X1.5) per the celeration period. For example, X1.5 per week indicates an increase of 50% in one week, whereas /1.5 indicates a decrease of 33% per week. Level typically refers to a measure of central tendency (e.g., mean, median, mode, etc.); for the purposes of this study, we used the geometric mean. Lastly, bounce refers to the variability present within the data set. One can quantify the degree of bounce by examining the proportional distance between the highest and lowest data points.

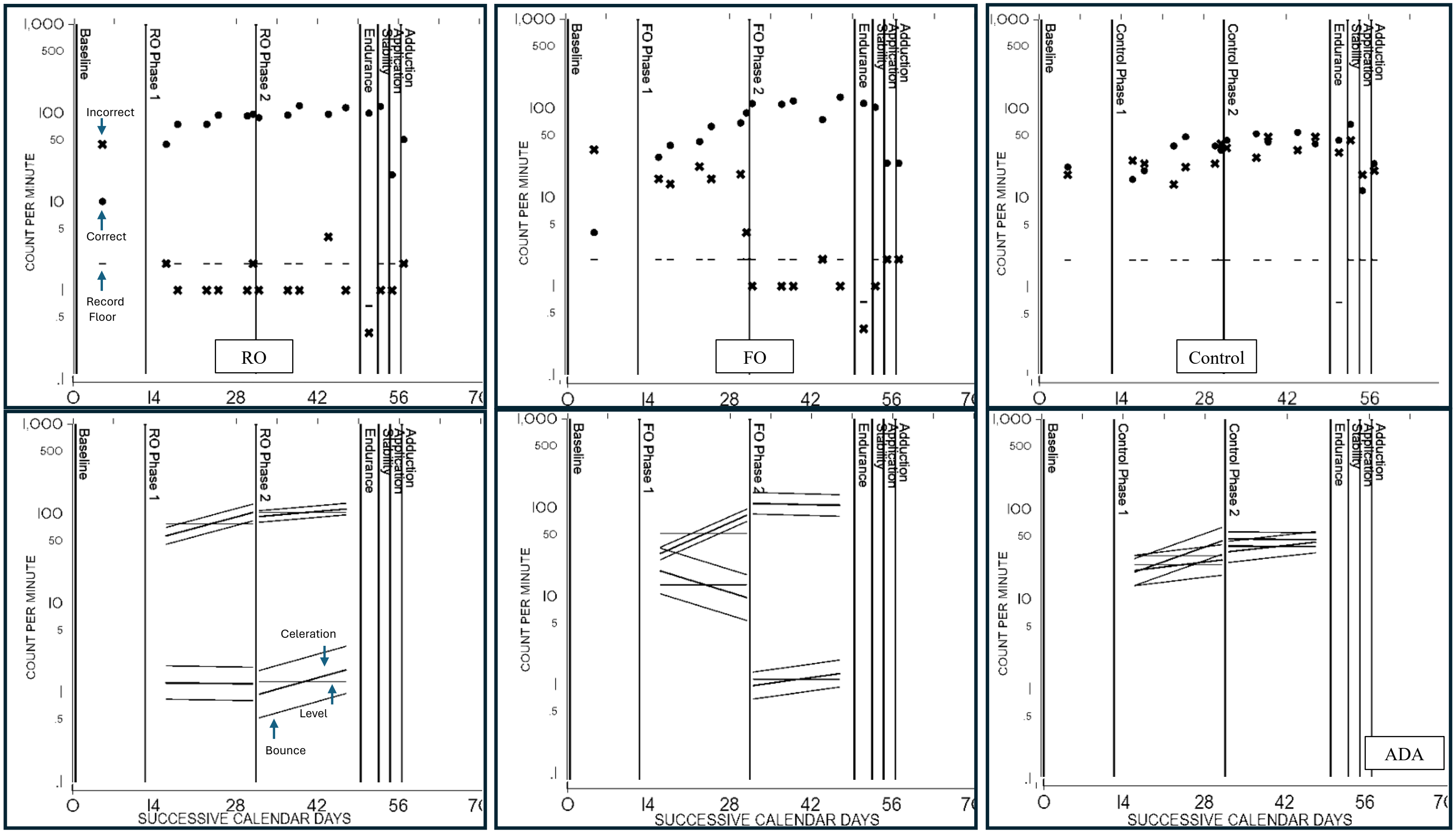
Accuracy

Overall, participants acquired accuracy of responding (zero incorrect responses) in fewer sessions in the RO condition (Mdn = 2; range, 7–2) than in the FO condition (Mdn = 7; range, 7–5), as displayed in Figures 1–5 and Table 2. Participants acquired accuracy more rapidly in both the RO condition and the FO condition than in the control condition. The control condition produced low levels of accuracy throughout both phases for all participants. All participants reached accuracy first under the RO condition.

Table 2. Number of Sessions to Accuracy.

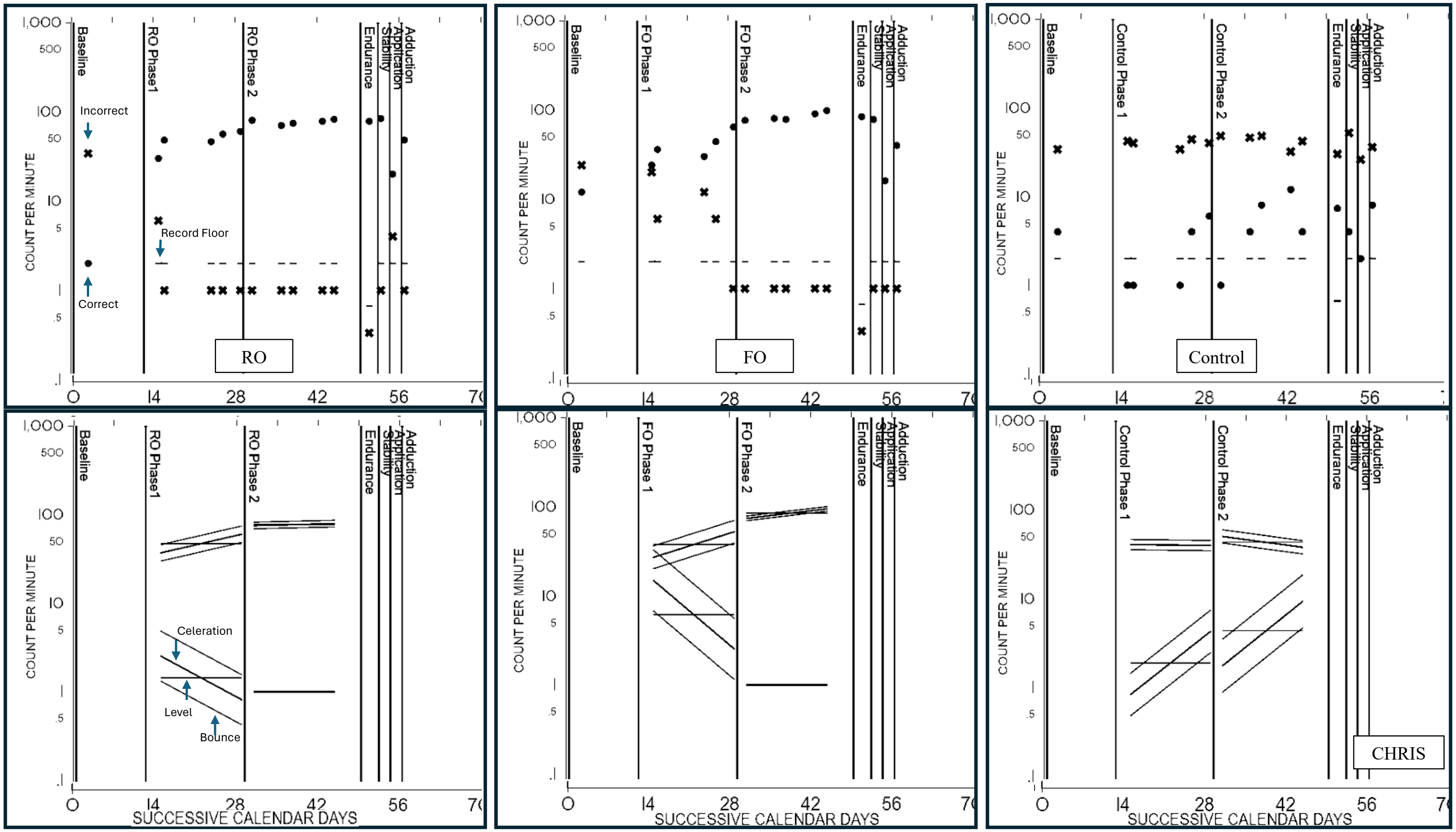
Participant	FO	RO	Control
Ada	7	2	N/A
Sheva	5	2	N/A
Chris	5	2	N/A
Leon	8	7	N/A
Jill	7	2	N/A

Figure 1. Results for Ada in the FO, RO, and Control Conditions.



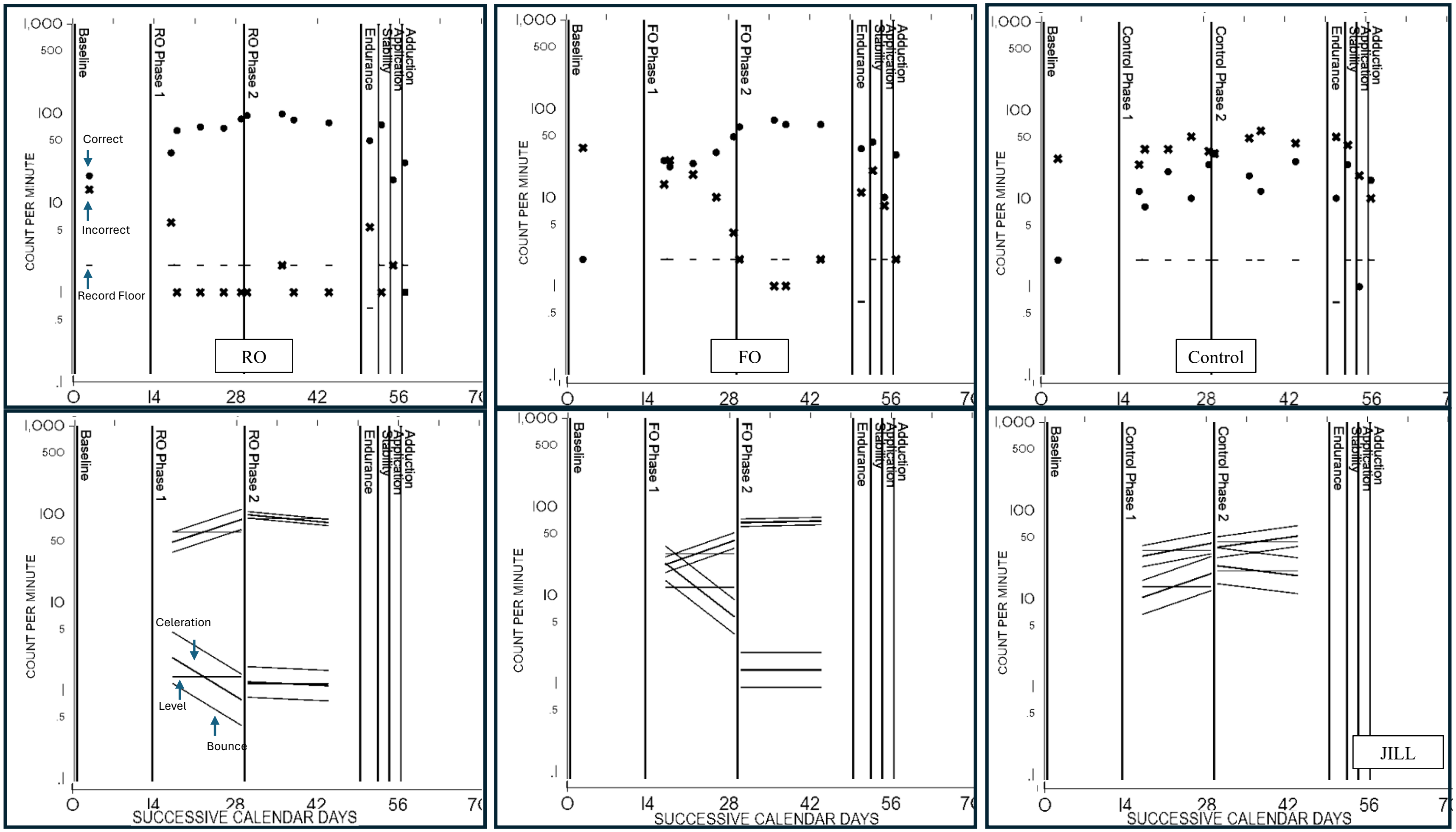
Note. This figure demonstrates the results for Ada. The top row displays the raw data, and the bottom row displays the data in aggregate form. The FO condition targeted Mandarin numerals, the RO conditions targeted Arabic numerals, and the control condition targeted Hindi numerals.

Figure 2. Results for Chris in the FO, RO, and Control Conditions.



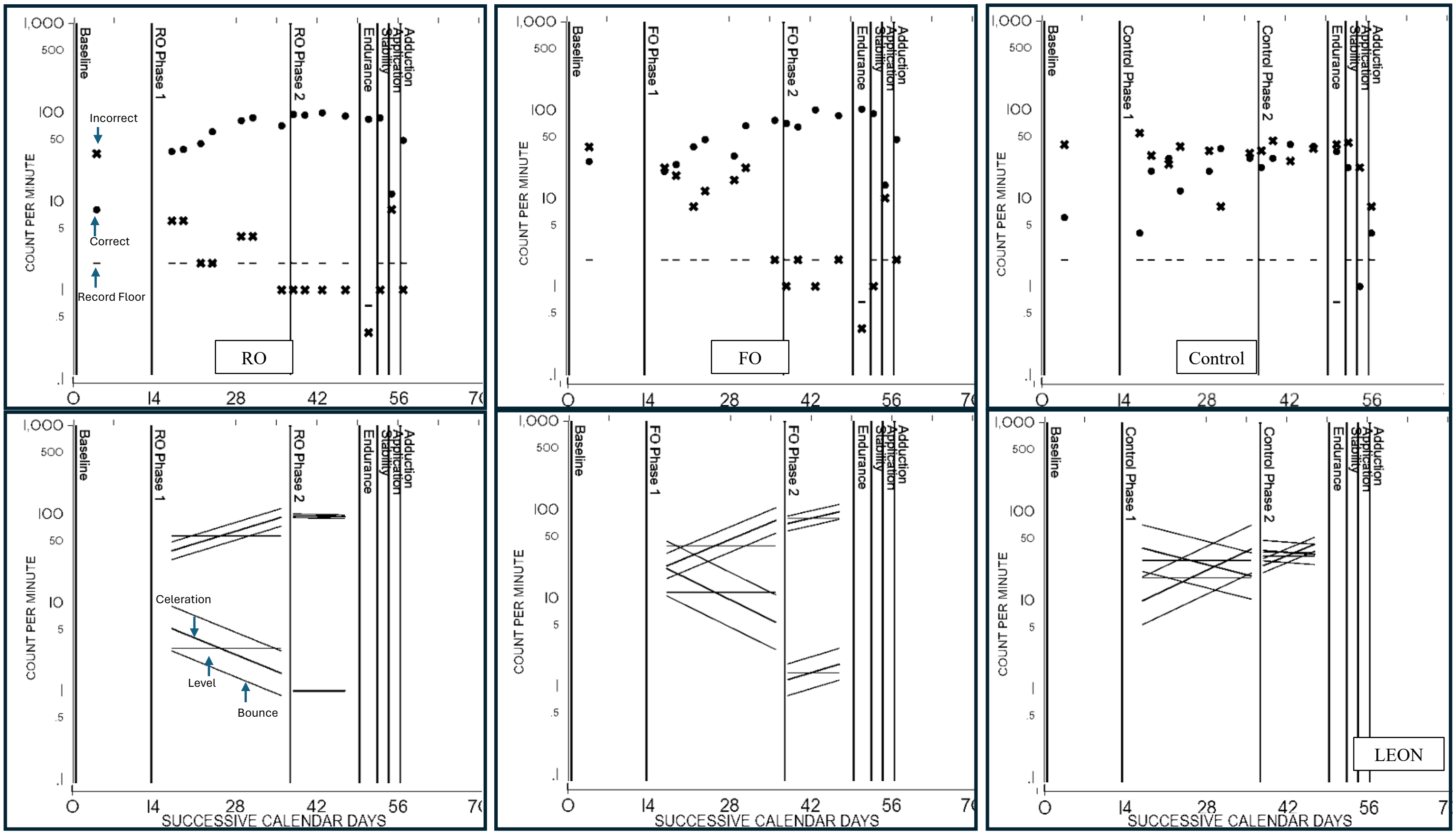
Note. This figure demonstrates the results for Chris. The top row displays the raw data, and the bottom row displays the data in aggregate form. The FO condition targeted Mandarin numerals, the RO conditions targeted Hindi numerals, and the control condition targeted Arabic numerals.

Figure 3. Results for Jill in the FO, RO, and Control Conditions.



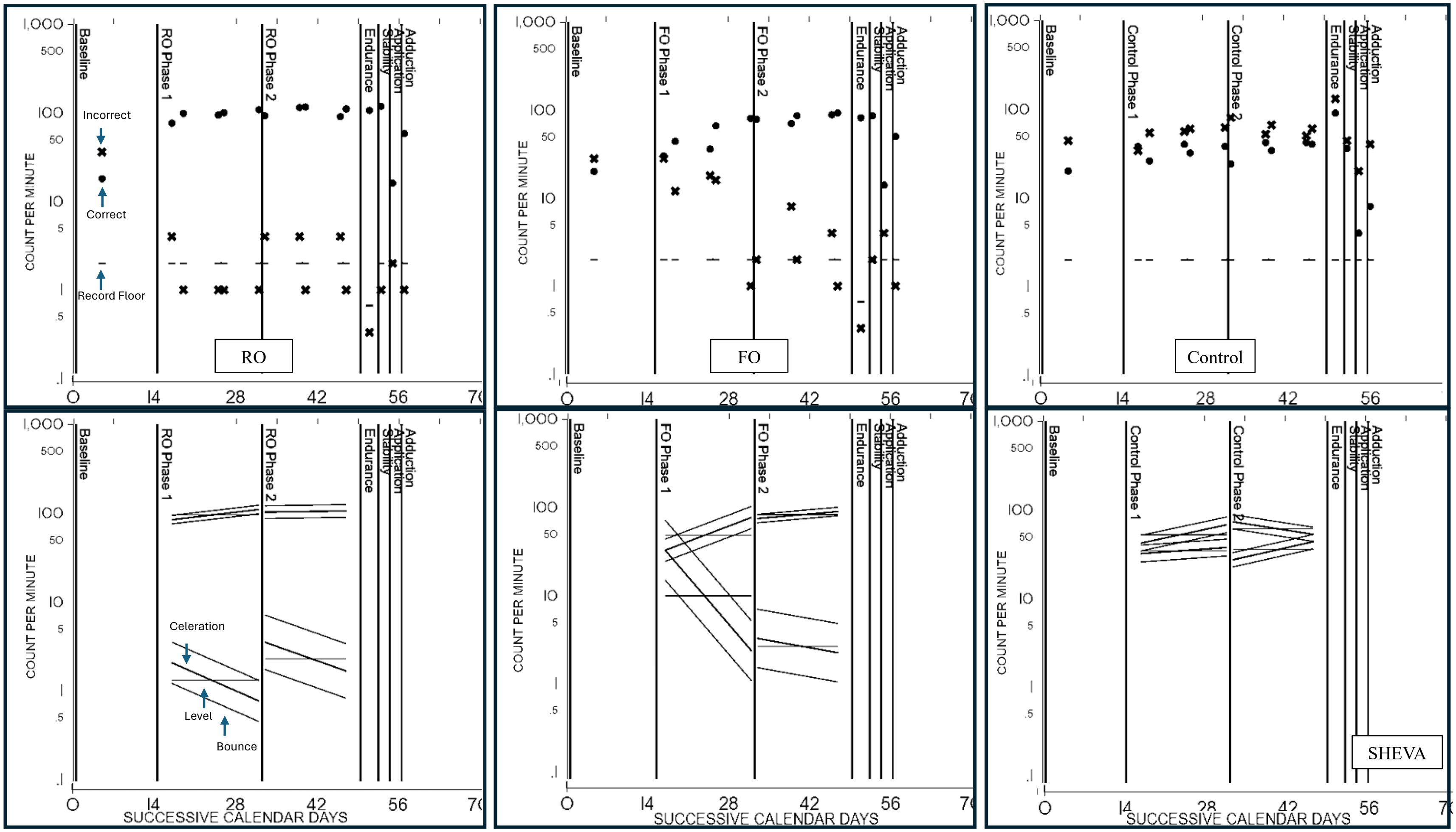
Note. This figure demonstrates the results for Jill. The top row displays the raw data, and the bottom row displays the data in aggregate form. The FO condition targeted Arabic numerals, the RO conditions targeted Mandarin numerals, and the control condition targeted Hindi numerals.

Figure 4. Results for Leon in the FO, RO, and Control Conditions.



Note. This figure demonstrates the results for Leon. The top row displays the raw data, and the bottom row displays the data in aggregate form. The FO condition targeted Hindi numerals, the RO conditions targeted Arabic numerals, and the control condition targeted Mandarin numerals.

Figure 5. Results for Sheva in the FO, RO, and Control Conditions.



Note. This figure demonstrates the results for Sheva. The top row displays the raw data, and the bottom row displays the data in aggregate form. The FO condition targeted Hindi numerals, the RO conditions targeted Arabic numerals, and the control condition targeted Mandarin numerals.

Level

Level represents a measure of central tendency throughout a condition. Overall, participants reached higher average levels of correct responding in the RO condition (Mdn = 62; range, 95– 47) than the FO (Mdn = 39; range 50–29) and control (Mdn= 18, range 34–1.9) in phase 1 and in phase 2 (, RO Mdn= 93, range 108– 77, FO Mdn= 82, range 102–67, control Mdn= 31, range 46– 4.3). As displayed in Table 3, Chris acquired a higher average level of correct responding in the FO condition than in the RO condition in phase 2, whereas all other participants acquired a higher average level of responding in the RO condition. Incorrect responses also occurred at a higher rate in the FO condition (Mdn = 12; range, 13–6.1) than in the RO condition (Mdn = 1.4; range, 3–1.3) and remained highest in the control condition (Mdn = 35; range, 52–24) in phase 1. In phase 2, the average incorrect responses occurred at a similar level in the FO (Mdn =1.4; range, 2.6–1) and RO (Mdn. = 1.2; range, 2.3– 1) conditions but remained high in the control condition (Mdn = 43; range, 61–34).

Table 3. Mean Level of Correct and Incorrect Responses.

Phase 1						
	Correct Responses			Incorrect Responses		
Participant	FO	RO	Control	FO	RO	Control
Ada	50	76	30	13	1.3	24
Sheva	48	95	34	10	1.3	52
Chris	37	47	1.9	6.1	1.4	40
Leon	39	56	18	12	3	28
Jill	29	62	14	12	1.4	35
Phase 2						
Ada	102	108	46	1	1.3	38
Sheva	82	104	36	2.6	2.3	61
Chris	84	77	4.3	1	1	43
Leon	79	93	31	1.4	1	34
Jill	67	88	21	1.4	1.2	44

Note. The table displays the geometric mean level of correct and incorrect responses per minute for each participant in each condition for each phase.

Celeration

Overall, in phase 1, correct responses grew with a moderate celeration in the RO (Mdn= X1.3, range X1.4– X1.1), FO (Mdn= X1.5, range X1.6– X1.4), and control (Mdn= X1.5, range X1.1– X2.3) as displayed in Figures 1–5 and Table 4. In the RO condition, a moderate value of deceleration of incorrect responses occurred since incorrect responses started and remained low (Mdn =1.6; range, /1.9– X1), whereas in the FO condition, incorrect responses decelerated over time, (Mdn = /2.3; range, /1.4– /3.4). In phase 2, insignificant accelerations occurred in the RO (Mdn = X1.1; range, X1.1– /1.1) and FO conditions (Mdn= X1; range, X1– X1.3) since the rate of correct responses reached high levels in the previous phase. The rate of correct responses continued to accelerate in phase 2 of the control condition for Sheva (X1.3), Chris (X2.3), and Leon (X1.5), whereas Ada (X1) and Jill's (X1.1) celerations remained relatively flat. Incorrect responses in phase 2 of the control condition also remained relatively flat (Mdn = /1.1; range, X1.2– /1.2).

Table 4. Celeration Values.

	Phase 1					
	Correct Responses			Incorrect Responses		
Participant	FO	RO	Control	FO	RO	Control
Ada	X 1.6	X1.3	X1.5	/1.4	X1	X1.1
Sheva	X1.5	X1.1	X1.1	/3.4	/1.6	X1.3
Chris	X1.4	X1.3	X2.3	/2.4	/1.8	/1
Leon	X1.6	X1.4	X1.6	/1.7	/1.5	/1.3
Jill	X1.4	X1.4	X1.4	/2.3	/1.9	X1.2
	Phase 2					
Ada	X1	X1.1	X1	X1.2	X1.3	X1.1
Sheva	X1.1	X1	X1.3	/1.2	/1.5	/1.2
Chris	X1.1	X1.02	X2.3	X1	X1	/1.2
Leon	X1.3	X1	X1.5	X1.4	X1	/1.1
Jill	X1	/1.1	X1.1	X1	X1	X1.2

Note. The table displays the quantitative growth (X) or decay (/) per week for correct and incorrect responses for each participant in each condition and in each phase.

Level & Celeration Multipliers

The level multiplier describes the exponential difference between the average level of responses per minute between conditions (Kubina, 2019). The multiplier expresses change as either 'X[value]' indicating a level increase from the reference condition to the subsequent condition, or '/[value]' indicating a decrease from a reference condition to the subsequent condition. X1 indicates no change between conditions. In phase 1, correct responding in the RO and FO conditions improved from the control condition. Additionally, all participants improved in the average level of responding from the FO to the RO conditions. In phase 2, both FO and RO conditions improved from performance in the control condition. A minimal difference occurred between the FO and RO conditions. See Table 5 for the quantitative difference in the level of correct responses between conditions for each participant.

Differences in incorrect responses between conditions occurred independently of differences in correct responses between conditions. Across all participants in phase 1, a decrease in the level of incorrect responses occurred in the FO and RO conditions compared to the control condition. The average level of incorrect responses increased from the FO to the RO condition across all participants. Table 5 displays the quantitative values of the level change between conditions in both phases for all participants.

The celeration multiplier describes the exponential difference between the rate of growth or decay between conditions. A value of X1 celeration indicates no change in the rate of growth/decay between conditions, whereas X2 indicates a doubling in the rate of response per week between conditions. In phase 1, little difference in celeration occurred for correct responses. Table 5 displays celeration multiplier values for correct responses for each participant, in both phases in each condition. Overall, the celeration of correct responses in phase 1 increased from the RO and FO condition compared with the control condition.

Table 5. Quantitative Comparisons of Correct Responses Between Conditions.

	FO				RO			
Participant	Endurance	Stability	Application	Adduction	Endurance	Stability	Application	Adduction
Ada	113/0	104/0	25/2	24/2	98/0	122/0	20/0	50/2
Sheva	81/0	90/2	14/4	50/0	108/0	118/0	16/2	57/0
Chris	84/0	78/0	16/0	40/0	78/0	84/0	20/4	48/0
Leon	100/0	90/0	14/10	46/2	82/0	86/0	12/8	48/0
Jill	35/11	42/20	10/8	31/2	50/5	74/0	18/2	28/0

Note. This table displays the differences in level (average responding) and celeration (Speed of growth/decay) between each of the conditions, for each participant and each level.

MESAA Outcomes

Overall, the FO and RO conditions produced higher rates of correct responses and lower rates of incorrect responses than the control condition across all fluency outcome checks. Between the RO and FO conditions, fluency outcomes occurred at a comparable level across most checks for all participants, as displayed in Tables 6 and 7. Ada obtained a higher rate of correct responses in the adduction check during the RO condition than the FO condition (X2.1 level multiplier). Jill also obtained a higher rate of correct responses in the application check (X1.9 level multiplier), and the stability check (X1.8 level multiplier) in the RO condition than in the FO condition. No significant outcomes favoring the FO condition over the RO condition occurred.

Table 6. MESAA Per Minute Frequencies in FO and RO Conditions.

	Phase 1					
	Level Multiplier			Celeration Multiplier		
Participant	FO compared to RO	Control compared to FO	Control compared to RO	FO compared to RO	Control compared to FO	Control compared to RO
Ada	X1.53	X1.66	X2.53	/1.2	X1.1	/1.15
Sheva	X1.98	X1.39	X2.75	/1.32	X1.38	X1.04
Chris	X1.25	X19.78	X24.72	/1.09	/1.65	/1.78
Leon	X1.45	X2.18	X3.17	/1.12	/1.05	/1.19
Jill	X2.14	X2.15	X4.60	/1.02	/1.01	/1.01
	Phase 2					
Ada	X1.06	X2.34	X2.2	X1.1	X1	X1.1
Sheva	X1.26	X2.3	X2.91	/1.07	X1.15	/1.25
Chris	/1.10	X19.40	X17.71	/1.11	/2.04	/2.26
Leon	X1.19	X2.53	X3.00	/1.30	/1.21	/1.57
Jill	X1.32	X3.24	X4.28	/1.13	X1.16	/1.03

Note. This table displays the correct and incorrect (c/i) results in checks for Endurance, Stability, Application and Adduction (RESAA).

Table 7. MESAA Per Minute Frequencies in the Control Conditions.

	Control			
Participant	Endurance	Stability	Application	Adduction
Ada	45/30	67/45	12/20	25/20
Sheva	88/126	40/46	4/20	8/40
Chris	8/30	4/52	2/26	8/40
Leon	34/40	22/40	0/25	4/8
Jill	10/50	24/40	0/18	15/10

Note. The table displays correct and incorrect (c/i) results in checks for Endurance, Stability, Application, and Adduction in the control condition.

Individual Analysis

The above results include a description of the effects of the three conditions (i.e., FO, RO, and control) in aggregate form for all five participants. In order to provide a model for how to analyze the results and demonstrate experimental control when using an AATD, we provide a detailed description for one participant, Ada. Figure 1 displays the data for Ada. During baseline, Ada produced a frequency of correct responses of 25 per min and 1 incorrect per min during the RO condition, 12 correct per min and 1 incorrect per min during the FO, and 12 correct per min and 10 incorrect per min for the control target set.

Level. During phase 1 of the RO condition, Ada had an average of 76 correct responses per min and 1.3 incorrect responses per min. For the FO condition in phase 1, Ada had an average frequency of 50 correct responses per min and 13 incorrect responses per min. Lastly, for the control condition, Ada had an average of 30 correct per min and 24 incorrect responses per min during phase 1. To meet the accuracy criteria, it took Ada 2 sessions during the RO condition, and 7 sessions during the FO condition. Ada did not meet the accuracy criteria for the control condition.

During phase 2, Ada produced 108 correct responses per min and 1.3 incorrect responses per min during the RO condition. In the FO condition, Ada had 102 correct responses per min and 1 incorrect response per min. For the control condition, Ada's average correct responses remained low with 46 correct responses per min and the average rate of incorrect responses remained high with 38 incorrect responses per min.

Celeration. During the RO condition for phase 1, Ada's correct responses grew by X1.3 (30% weekly growth) and celeration for incorrect responses came to X1 (0% weekly growth/decay, no trend). During the FO condition, her correct responding grew by X1.6 (60% weekly growth) and incorrect responding decelerated by /1.4 (29% weekly decay). Celeration during the control condition had a slight celeration of X1.1 (10% weekly growth).

During phase 2, Ada's frequency of correct and incorrect responses grew by X1.1 (10% weekly growth) and X1.3 (30% weekly growth), respectively. In the FO condition, the frequency of correct responses remained stable during with a X1 celeration (0% weekly growth, no trend) and a X1.2 (20% weekly growth) celeration for incorrect responses. For the control condition, celeration for correct responses remained stable with a celeration of X1 and incorrect responses came to X1.1 (10% weekly growth).

*Level Multiplier***.** When comparing performance across the conditions for phase one, Ada had a X1.53 (53% more) higher average rate of correct responding during the RO condition when compared to the FO condition. When comparing the FO condition to the control condition, Ada had a X1.66 (66% more) higher average rate of correct response. When compared to the control, RO produced X2.53 (150%) more correct responses. During phase 2, Ada had X1.06 (6%) more correct responses per min during the RO condition compared to the FO condition. When compared to the control condition, Ada produced a level multiplier of X2.34 (134% more compared to control) for the FO condition and X2.2 (120% more compared to control) for the RO condition.

*Celeration Multiplier***.** During phase 1, Ada had a celeration multiplier of /1.2 indicating that her frequency of correct responses per min across time occurred slower during the RO condition than in the FO condition. When comparing the control condition to the FO condition the celeration multiplier came to X1.1, indicating that correct responses occurred 1.1 times faster in the FO condition than the control condition. The celeration multiplier analysis for comparing the control condition to the RO condition came to /1.09 indicating a slower change in rate of responding during the RO condition. In phase 2, Ada's correct responding grew 1.1 times more during the RO condition than in the FO condition. When comparing the control condition to the FO condition, the celeration multiplier analysis found no difference in the speed in which behavior changed (X1). For the control to RO comparison, correct responses during the control condition occurred X1.1 faster than in the RO condition.

Discussion

Both the RO and FO conditions produced effective skill acquisition for all learners. Although no participant reached the recommended frequency aims indicative of fluent performance, overall, the participants achieved performances during the stability and endurance checks in both RO and FO conditions comparable to their performances in phase 2. Performances in the application checks and adduction checks varied. Given the robust literature base for fluency outcomes after attaining higher frequency aims, the need to abruptly cease the study likely affected any differentiation between the level of responding achieved and MESAA outcomes between conditions. Likewise, not reaching frequency aims in either condition limited the ability to detect an effect of yoking responses from the FO to the RO condition.

However, the clear difference in the rate of acquisition of accurate responses in the RO condition compared with the FO and control conditions remains of note. Leon reached accuracy in the RO condition on the seventh session and in the eighth under the FO conditions, whereas all other participants reached 100% accuracy within two sessions in the RO condition compared to five to eight sessions before reaching accuracy in the FO condition. During the acquisition phased of learning, imposing some restriction on the rate of response as we did with the RO condition, resulted in the most efficient rate of acquisition between the three conditions. These results may provide some evidence indicating the benefit of an "accuracy first" approach to teaching a new skill as all participants, on average, acquired the skill under RO conditions roughly twice as fast.

Several pragmatic benefits may occur when one trains a new behavior to accuracy under restricted conditions prior to increases the frequencies of the response under freer conditions. First, fewer errors consistently occurred under RO conditions during the acquisition of these novel discriminations. Thus, higher relative rates of reinforcement for correct responses compared to incorrect responses occurred under RO conditions compared to early FO sessions. Discoveries from the Experimental Analyses of Behavior laboratories suggest other benefits of reducing errors during skill acquisition. Faster skill acquisition limits exposure to extinction, the aversiveness of the contingency (Knutson, 1970), and the emotional behavior that extinction can produce. Extinction-induced emotional behavior may interfere with operant behavior necessary in contacting new contingencies of reinforcement, delaying the acquisition of new discriminations (Pierce & Cheney, 2017). In addition to faster acquisition, learning without errors may produce more resistance to extinction once acquired. Marsh and Johnson (1968) found that errorless training produced lasting effects such that pigeons would not respond to the Sdelta even after researchers reversed S^D and Sdelta roles. These studies suggest that preventing errors not only allowed for more efficient acquisition of a skill; they also produced robust discrimination.

Second, by initially imposing some restriction in the form of a longer ITI, the learner may have time to physically attend to the features of the presented stimuli better than under FO conditions where researchers instructed them to go as fast as they could. This additional time may allow for longer contact between the presented stimulus and sensory neurons, thus promoting quicker learning. Additionally, for verbal learners, this additional time may also prompt covert verbal behavior that they can use to mediate the response in the future. For example, when presented with the Arabic numeral /٣ / that represents the number three, the learner has the opportunity to say things like, “Oh that is three. I can see a sideways three on the top. That makes sense.” When they experience the same stimulus again, that covert behavior may help mediate the response, such as when a learner says, “There’s the sideways three. This is the number three.” Future research can begin to investigate these variables to identify the extent to which they aid in learning. To control for stimulus and contact time, researchers can yolk the time the learner spent viewing each stimulus under the RO condition to the time spent viewing the stimulus under FO conditions. Additionally, to account for covert mediators, previous researchers have required participants to engage in overt verbal responses not related to the task as an incompatible behavior to covert response (e.g., Ratkos et al., 2016).

Future researchers should also account for some of the limitations to the current investigation. Given the abrupt end to the study, researchers should continue with the study until participants reach their pre-determined frequency aims that researchers have investigated as predictors of fluency. Additionally, the teaching stimuli under FO and RO conditions differed in size, with RO procedures using larger images of the stimuli. We did this because to have multiple stimuli on the FO sheets, we had to shrink the size of the stimuli to fit on one sheet. This may have inadvertently allowed participants to better attend to features of the stimuli under RO conditions compared to FO conditions. Similarly, the probe sheets used for measure of the dependent variable resembled the practice sheets used during the FO conditions. This represents an intra-experimental variable that may have influenced the results, given the stark difference of the practice format used in the RO conditions.

Providing the most efficient and effective teaching paradigm greatly benefits learners and teachers alike. The comparative overall effects of outcomes between teaching conditions still warrant exploration, including exploring when to “free up” the operant for optimum learning. This may guide practicing behavior analysts as to when they should use each procedure. Johnson et al. (2020) present three phases of learning used in the Morningside Model of Generative Instruction (MMGI): (1) Acquisition, (2) Fluency, and (3) Application. MMGI incorporates the use of RO teaching paradigms during the acquisition stage of learning with technologies based on direct instruction. They also use FO procedures during the practice stage of learning, to build previously established skills to fluent performances. Give the robust data on the effectiveness of MMGI coupled with the findings of the current study, it appears that it is not an argument of whether to use RO or FO teaching methodologies, but a question of *when* to use each arrangement.

References

Aravamudhan, S., & Awasthi, S. (2021). The use of prompts and precision teaching to address speech sound disorders in a 17-year-old girl with autism. *Behavior Analysis in Practice*, *14*(3), 644–659. https://doi.org/10.1007/s40617-020-00470-7

Binder, C. (1993). Behavioral fluency: A new paradigm. *Educational Technology, 33*(10), 8–14. https://www.jstor.org/stable/44428106

Bulla, A. J. (2023) T*owards an Understanding of the Continuum of Operant Freedom: Some Recent Basic and Conceptual Work* [conference presentation]. The 36th Annual Conference of the Standard Celeration Society, St. Petersburg, Florida.

Cariveau, T., & Fetzner, D. (2022). Experimental control in the adapted alternating treatments design: A review of procedures and outcomes. *Behavioral Interventions*, *37*(3), 805-818. https://doi.org/10.1002/bin.1865

Cariveau, T., Irwin Helvey, C., Moseley, T. K., & Hester, J. (2022). Equating and assigning targets in the adapted alternating treatments design: Review of special education journals, Remedial and Special Education, 43(1). 58-71. https://doi.org/10.1177/0741932521996071

Doughty, S. S., Chase, P. N., & O'Shields, E. M. (2004). Effects of rate building on fluent performance: A review and commentary. *The Behavior Analyst*, *27*, 7–23. https://doi.org/10.1007/BF03392086

Evans, A. L., Bulla, A. J., & Kieta, A. R. (2021). The precision teaching system: a synthesized definition, concept analysis, and process. *Behavior Analysis in Practice*, *14*(3), 559 –576. https://doi.org/10.1007/s40617-020-00502-2

Fabrizio. M.A, & Moors, A.L., (2003) Evaluating mastery: Measuring instructional outcomes for children with autism. *European Journal of Behavior Analysis, 4*(1 –2), 23 –36. https://doi.org/10.1080/15021149.2003.11434213

Ghezzi, P.M. (2007). Discrete trials teaching. *Psychology in the Schools, 44*(7), 667 – 679. https://doi.org/10.1002/pits.20256

Gist, C. & Bulla, A. J. (2022). A systematic review of frequency building and precision teaching with school-aged children. *Journal of Behavioral Education, 31,* 43 –68. https://dio.org/10.1007/s10864-020-09404-3

Johnson, K., & Street, E. M. (2012). From the laboratory to the field and back again: Morningside Academy's 32 years of improving students' academic performance. *The Behavior Analyst Today*, *13*(1), 20–40. https://doi.org/10.1037/h0100715

Johnson, K., Street, E.M., Kieta, A.R., Robbins, J.K., (2020). *The Morningside model of generative instruction. Building a bridge between skills and inquiry teaching.* Sloan Publishing.

Knutson, J. (1970). Aggression during the fixed-ratio and extinction components of a multiple schedule of reinforcement. *Journal of the Experimental Analysis of Behavior, 13* 221 –231. https://doi.org/10.1901/jeab.1970.13-221

Kubina, R. M. (2019). *The precision teaching implementation manual*. Greatness Achieved.

Lindsley, O. R. (1996). The four free-operant freedoms. *The Behavior Analyst, 19*(2), 199–210. https://doi.org/10.1007/BF03393164

Marsh, G., & Johnson, R. (1968). Discrimination reversal learning without "errors." *Psychonomic Science, 10*, 261–262. https://doi.org/10.3758/BF03331510

McTiernan, A., Holloway, J., Healy, O., Hogan, M. (2016). A randomized controlled trial of the Morningside math facts curriculum on fluency, stability, endurance and application outcomes. *Journal of Behavioral Education*, *25*(1), 49 – 68. https://doi.org/10.1007/s10864-015-9227-y

Nopprapun, M., & Holloway, J., (2014). A comparison of fluency training and discrete trail instruction to teach letter sounds to children with ASD: Acquisition and learning outcomes. *Research in Autism Spectrum Disorders, 8,* 788 – 802. http://dx.doi.org/10.1016/j.rasd.2014.03.015

Pennypacker, H.S., Gutierrez Jr., A., & Lindsley, O.R. (2003). *Handbook of the standard celeration chart*. Cambridge Center for Behavioral Studies.

Pierce, W. D. & Cheney, C. D. (2017). *Behavior analysis and learning: A biobehavioral approach* (6th ed.). Psychology Press. https://doi.org/10.4324/9780203441817

Ratkos, T., Frieder, J.E. & Poling, A. (2016) Accurate delayed matching-to-sample responding without rehearsal: An unintentional demonstration with children. *The Analysis of Verbal Behavior, 32*, 69–77. https://doi.org/10.1007/s40616- 016-0052-8

Richling, S. M., Williams, W. L., & Carr, J. E. (2019). The effects of different mastery criteria on the skill maintenance of children with developmental disabilities. *Journal of Applied Behavior Analysis*, 52(3), 701–717. https://doi.org/10.1002/jaba.580

Stocker, J.D., Schwartz, R., Kubina, R.M., Kostewicz, D., & Kozloff, M. (2019). Behavioral fluency and mathematics intervention research: A review of the last 20 years. *Behavioral Interventions, 34*(1), 102 –117. https://doi.org/10.1002/bin.1649

Wolery, M., Gast, D. L., & Ledford, J. R. (2014). Comparison designs. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (pp. 297–345). Routledge.

Wolery, M., Gast, D. L., & Ledford, J. R. (2018). Comparative designs. In J. R. Ledford & D. L. Gast (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (3rd ed.). Routledge.