The Effects of Email Performance-Based Feedback on Teacher Candidate Implementation of a Multiple Stimulus Without Replacement Preference Assessment

Emma S. Sipila-Thomas1 · Matthew T. Brodhead1 · Ashley N. Walker1

Accepted: 4 October 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
The purpose of this study was to evaluate the effects of performance-based feedback delivered via email on teacher candidates’ implementation of a multiple stimulus without replacement (MSWO) preference assessment with an adult confederate. Six participants enrolled in a special education teacher preparation program were recruited to participate. After exposure to an introductory MSWO training video, participant fidelity of MSWO administration was measured. Following baseline, performance feedback was delivered via e-mail to each participant after the completion of each research session. Results indicate that performance feedback delivered via e-mail improved fidelity of implementation for all six participants. Implications and future directions are discussed.

Keywords E-mail · Multiple stimulus without replacement · Performance-based feedback · Telehealth

Introduction
Performance feedback is a critical component of professional development (Barton et al., 2020; Miltenberger, 2012). Performance feedback involves the use of data, derived from an observation occurring during supervision, to inform the delivery of feedback in order to change and sustain the individual’s behavior (Barton et al., 2016, 2020; Hemmeter et al., 2011; Novak et al., 2019). Within school settings,
researchers found performance feedback to increase procedural fidelity and maintain the teacher’s use of effective practices, which in turn increased the quality of instruction provided and improves child learning outcomes (Barton et al., 2020; Schles & Robertson, 2019). Without additional support (e.g., performance feedback), new and returning teachers may implement evidence-based practices with low or variable levels of fidelity and negatively impact child learning outcomes for students with disabilities (Schles & Robertson, 2019).

Performance feedback can be delivered in many forms, such as in verbal or written formats, and can be provided during or after a supervisory observation of the target individual implementing an intervention or engaging in an activity (Barton & Wolery, 2007). In addition, performance feedback can be delivered in a variety of modalities such as bug-in-ear as well as visible counters, public wall postings, and personal interactions (Coogle et al., 2016, 2017; Warrilow et al., 2020). As technology has evolved and become more available, so have modalities for delivering performance feedback, such as computer displays, text messages, video conferencing, social media communications, and emails (e.g., Barton & Wolery, 2007; Hemmeter et al., 2011; Krick Oborn & Johnson, 2015; Zhu et al., 2021).

One technology-based modality that has been shown to have several advantages for delivering performance feedback is email (Gorton et al., 2021). Email feedback allows supervisors to save time by sending an email following the observation rather than scheduling a time to meet in person to review that feedback (Warrilow et al., 2020). Related, the observer is able to send the email feedback immediately to the individual after the observation is completed, without interrupting the individual and the activity they are engaging in (e.g., implementing an intervention with a child; Barton & Wolery, 2007; Gorton et al., 2021; Zhu et al., 2021). Second, verbal and some forms of written feedback (e.g., handwritten notes) may be seen as obsolete and ineffective by the individuals receiving the feedback, whereas wireless communication forms of feedback (e.g., email, video, bug-in-ear) are seen as more current or up-to-date (Barton et al., 2020; Gomez et al., 2021; Zhu et al., 2021) and may be seen as more socially acceptable (Barton & Wolery, 2007). Additionally, using emails results in an electronic record of the feedback provided (Zhu et al., 2021), which can be reviewed more than once and can be used for future performance reviews. Overall, email feedback may be a strategy for interacting with individuals quickly and more efficiently (Barton et al., 2016; Warrilow et al., 2020).

A limitation of previous email feedback research is that the studies did not isolate the effects of email feedback, because the email feedback was always provided in conjunction with a training (Artman-Meecker & Hemmeter, 2013; Gomez et al., 2021; Gorton et al., 2021; Hemmeter et al., 2011; Martinez Cueto et al., 2021). Researchers then sought to isolate and evaluate the effects of email feedback alone (Barton et al., 2016, 2020) and also compared email feedback alone to other forms of feedback such as immediate bug-in-ear feedback (Coogle et al., 2020) and videoconference feedback (Zhu et al., 2021). Overall, the studies found that email feedback alone was effective in increasing teachers’ behaviors.

Though the aforementioned studies evaluated providing performance feedback via email, there were several limitations to the studies that the present study hopes to address. First, within the recent studies evaluating email feedback, researchers did
not control for the amount or quality of the performance feedback provided in each email. This lack of experimental control presents problems because researchers have found that more frequent and specific feedback produces more significant changes in performance (Park et al., 2019; Sleiman et al., 2020). It is unclear if variation in amount or quality of performance feedback in previously published research differentially affected participant outcomes.

Second, several of the aforementioned studies included a training component between the baseline condition and the intervention condition (i.e., Artman-Meecker & Hemmeter, 2013; Gomez et al., 2021; Gorton et al., 2021; Hemmeter et al., 2011; Martinez Cueto et al., 2021). Thus, it is unclear if improvements in instructional behaviors were a result of email feedback alone, training alone, or a combination of the email feedback and the training.

Third, the previous literature did not control for variation in the instructional environment. Specifically, the studies reviewed above involved classroom settings where teachers engaged with students. Opportunities for teacher participants to respond, then, were at least particularly affected by student behavior in those settings, and it is unclear to what extent participant behavior was at least partially affected by student behavior (e.g., frequency of teacher descriptive praise is dependent on the frequency of student engagement).

In summary, the purpose of the present study is to extend previous research on performance feedback delivered via email by evaluating the effects of email feedback alone on teacher candidates’ implementation of a multiple stimulus without replacement (MSWO) preference assessment. An MSWO preference assessment was selected to assess the effects of email feedback because an MSWO is a common preference assessment used within the field of behavior analysis that is well described in the literature and consists of discrete steps that can be easily measured (e.g., represents remaining items and rotates by taking the item at the left end of the line and moving to the right end, then shifting the other items so they are equally spaced on the table), and the duration to conduct an MSWO is shorter in comparison to other preference assessments (e.g., paired-choice preference assessments; Graff & Karsten, 2012; Kang et al., 2013; Leaf et al., 2015; Lill et al., 2021). Therefore, researchers evaluated the following research question: Given email feedback, to what extent did teacher candidates’ implement an MSWO preference assessment with fidelity?

**Method**

**Participants**

Participants were recruited from three undergraduate courses offered through a special education program at a Midwestern University. A total of six participants (all identified as female) were recruited and agreed to participate in the study. Participants were between 20 and 23 years of age. Participants were included in the study if they: (a) were enrolled at the university at the start of the study, (b) had no prior experience implementing an MSWO preference assessment, (c) had access to an
email service, (d) had access to a computer with audio and video capabilities, and (d) were willing to read and respond to emails daily. Participants were screened for the inclusion criteria through a questionnaire prior to the start of the study. In addition, each participant was required to respond to three non-study-related emails sent over three days prior to the start of the study.

Riley was a 20-year-old female who identified as White and not of Latinx, Hispanic, or Spanish origin. Riley had been enrolled at the university for less than one year, was a transfer student, and was majoring in Special Education.

Olivia was a 23-year-old female who identified as Asian and not of Latinx, Hispanic, or Spanish origin. Olivia had been enrolled at the university for four years and was majoring in Special Education.

Ava was a 20-year-old female who identified as White and not of Latinx, Hispanic, or Spanish origin. Ava had been enrolled at the university for three years and was majoring in Special Education.

Layla was a 20-year-old female who identified as White and not of Latinx, Hispanic, or Spanish origin. Layla had been enrolled at the university for two years, was a transfer student, and was majoring in Special Education.

Ellie was a 21-year-old female who identified as White and not of Latinx, Hispanic, or Spanish origin. Ellie had been enrolled at the university for four years and was majoring in Special Education and Elementary Education.

Kennedy was a 20-year-old female who identified as Black or African American and Hispanic, and of Latinx, Hispanic, or Spanish origin. Kennedy had been enrolled at the university for three years and was majoring in Special Education.

Confederate

The primary researcher, a doctoral candidate who held a BCBA credential and had prior experience conducting research on preference assessments with children with ASD, served as the confederate throughout the duration of the study. The confederate played the role of a student (i.e., the learner) and engaged in specific responses during each research session (see Table 1 and see Confederate Response Data Sheet section for further description).

| Response After the Participant Says, “Pick One or Choose One” |
|---------------------------------------------------------------|
| Immediately select one stimulus from the array presented (one item) |
| Wait 7 s and then select one stimulus from the array presented (one item) |
| Immediately select two stimuli simultaneously (simultaneous) |
| Wait 7 s and then select two stimuli simultaneously (simultaneous) |
| Immediately select one stimulus and then select another stimulus from the array presented (sequential) |
| Wait 7 s and then select one stimulus and then select another stimulus from the array presented (sequential) |
| Do not select a stimulus from the array presented (no response) |
Primary Data Collector

One graduate student served as the primary data collector and measured the dependent variable (participant procedural fidelity) for all research sessions across all conditions for each participant. The graduate student was trained by the primary researcher on the data collection process prior to the start of the study. The training was conducted across two different days, for 1 h each day. During the training session, the primary researcher reviewed each step of the participant procedural fidelity data sheet (see Dependent Variable section for further description), virtually displayed the MSWO training video that the participants were presented (see Participant Materials section for further description) and practiced scoring a research session. Following the training sessions, both the primary researcher and the primary data collector scored three research sessions independently. If the primary data collector achieved 100% accuracy across all three research sessions, they were considered to pass the reliability checks and were assigned to score the remaining research sessions. The primary data collector achieved 100% accuracy across the three research sessions the first time.

Setting and Materials

Due to the COVID-19 pandemic, all research sessions, data collection, and feedback were conducted or provided remotely. For each research session, the participant and the primary researcher/confederate were present. Additionally, each research session was recorded for data collection purposes, and no data collection occurred during the session. Instead, research sessions were reviewed later by the primary data collector. The primary data collector reviewed and scored research sessions in the order in which they were recorded.

Participant Materials

Participant materials consisted of a computer with audio and video capabilities, an email service, a video conferencing software (e.g., Zoom), a pen or pencil, a calculator, a timer (e.g., phone), five leisure stimuli, and data sheets. The five leisure stimuli (e.g., ball, stapler) for the preference assessment were arbitrarily selected by each participant based on the stimuli they had available in their households for each research session. Twenty data sheets (see Supplementary Materials, Appendix A) for participants to use for data collection while conducting the MSWO preference assessment were mailed to the participant’s household or emailed to the participant (based on each participant’s preference) prior to the start of the study.

Primary Researcher/Confederate Materials

Primary researcher/confederate materials consisted of a computer with audio and video capabilities, a video conferencing software (e.g., Zoom), an email service,
a timer (e.g., digital stopwatch), and a second computer monitor. Additional materials consisted of an MSWO training video (presented once during the initial research meeting), an MSWO excerpt, a session script, a list of responses (here after referred to as the confederate response data sheet) to perform during each research session (see Table 1), and a pool of specific responses for each component of the email feedback. During each research session, the primary researcher/confederate used a second monitor to display the assigned confederate response data sheet, timer, and the session script. The primary monitor was used to display the video conferencing software in gallery view, to allow both the primary researcher/confederate and participant to be visible side by side.

**MSWO Training Video** The purpose of the MSWO training video was to approximate an “onboarding” or initial training a professional may receive on the topic and to provide basic information about how to implement an MSWO preference assessment, so that each participant would have baseline level knowledge of the MSWO prior to the start of the study. The MSWO training video was informed by an MSWO video created by researchers at Vanderbilt University (Chazin & Ledford, 2016) and by the primary researcher’s eight years of ABA experience and implementing preference assessments. The training video was created by the primary researcher and included a step-by-step guide of how to conduct an MSWO preference assessment within a remote context. When an MSWO preference assessment is conducted in-person, the learner may be able to engage with the physical item following a selection response. However, when an MSWO preference assessment is conducted within a remote context, the learner may not be able to engage with the physical item because they may not have access to it. Instead, the implementer may engage with the item and the learner may observe. The training video included specific steps that were tailored to meet the unique aspects of conducting an MSWO preference assessment within a remote context.

The video included a video model where the primary researcher played the role of the participant and a research assistant played the role of the learner (i.e., confederate), written text that indicated the target behavior for each step, and a voice over narration completed by the primary researcher. The training video (see Brodhead, 2022) was 21 min in length and was displayed virtually to the participants by the primary researcher one time prior to the start of the study (i.e., during the initial research meeting that occurred one week before each participant’s first research session). The participants did not have access to the training video outside of the initial research meeting.

**MSWO Excerpt** The MSWO excerpt was derived from a relevant peer-reviewed article (DeLeon & Iwata, 1996) and consisted of sections that outlined correct implementation of the MSWO preference assessment (see Supplementary Materials, Appendix B). The purpose of the MSWO excerpt was to provide the participants with a brief description of the MSWO procedures in order to mimic a description someone would find if they were to conduct an internet search or access notes on MSWO preference assessments. Additionally, this was an identical approach that previous preference assessment literature has used (e.g., O’Handley et al., 2021; Rosales et al., 2015).
The MSWO excerpt was displayed virtually to each participant at the beginning of each research session for up to 15 min. During the 15 min, the participants could read and reread the entire excerpt or specific sentences. For one participant, due to participant characteristics that cannot be disclosed for confidentiality reasons, the MSWO excerpt was displayed virtually to the participant and was read aloud to the participant by the primary researcher for up to 15 min. Similar to the other participants, the participant could ask the primary researcher to read or reread the entire excerpt or specific sentences aloud. At any time during the 15 min, the participants could indicate to the primary researcher that they were finished reviewing the excerpt and the primary researcher would stop displaying or reading aloud the excerpt. The participants did not have access to the excerpt outside of the research sessions.

**Session Script** The session script consisted of step-by-step instructions for the primary researcher to engage in during each research session. It included directions for the primary researcher to share the MSWO excerpt and how to respond if a participant asked a question which ensured consistency in researcher responses within and across participants. See Supplementary Materials, Appendix C for a complete description of the session script.

**Confederate Response Data Sheet** Each confederate response data sheet consisted of a sequence of responses for the confederate to engage in during the research session (see Supplementary Materials, Appendix D for an example). Specifically, each data sheet included 15 responses, one for each trial of the MSWO. There were a total of 10 variations in the confederate response data sheet.

The sequence of the responses on each confederate data sheet were randomly generated. There were seven different responses the confederate could engage in (see Table 1) throughout a research session. Given that each research session consisted of 15 total trials, some of the responses were included in the sequence more than once and some were included only one time. To ensure that each of the seven responses were included at least once within each of the confederate response data sheet variations, the primary researcher used a random number generator (Maple Tech International LLC, 2022) to identify a trial number for each potential response to occur in. Once all seven responses were included in the sequence, the primary researcher then used the random number generator to determine how many additional times each response would be included in the data sheet, with no response being repeated more than three times. Once the frequency of each response was determined, the primary researcher continued to use the random number generator to determine which trial each response would occur in until all trials had a response.

For each research session, the assigned confederate response data sheet was randomly selected using a random number generator (Maple Tech International LLC, 2022). Once a confederate response data sheet (e.g., assigned confederate response data sheet #8) was used for a participant, it was not replaced into the potential selections for future research sessions, until all 10 confederate data sheets had been used for that participant. Once all 10 confederate response data sheets had been used for that participant, all 10 had the opportunity to be selected again.
Primary Data Collector Materials

The primary data collector materials consisted of a pen or pencil, a timer, the video recording of the research session, an email service, and data sheets. The primary data collector used three different data sheets. The first data sheet was used to record the confederate’s stimulus selections, which was identical to the participant’s data sheet. The second data sheet was used to record participant’s adherence to the steps of the MSWO preference assessment, the dependent variable (i.e., procedural fidelity). Finally, the third data sheet was the assigned confederate response data sheet, so the primary data collector would know the sequence of responses the primary researcher/confederate engaged in during the research session.

Measurement

Dependent Variable

The primary dependent variable was fidelity of participant implementation (procedural fidelity). Procedural fidelity was the degree to which the participant implemented the MSWO as intended (Cooper et al., 2007; Gast & Ledford, 2014). The dependent variable comprised of component responses derived from a task analysis that depicted the instructional behaviors for the MSWO preference assessment (see Supplementary Materials, Appendix E). The task analysis was developed using the MSWO steps provided in DeLeon and Iwata (1996), DeLeon et al. (1997), and Sipila-Thomas et al. (2021) and modified for delivery in a remote context. The task analysis for the MSWO contained 24 total steps. Each step in the task analysis was coded as a whole occurrence (i.e., always occurred) or a whole nonoccurrence (i.e., never occurred). In order for a participant’s response to be scored as a whole occurrence, the participant had to correctly engage in the response multiple times (e.g., engaged in the response correctly seven out of seven times) throughout the research session. If at any point the participant did not engage in the response correctly (e.g., engaged in the response correctly five out of seven times), the entire response was scored as a whole nonoccurrence for that session. The percentages of occurrences for each session were calculated by dividing the total number of occurrences (steps implemented correct in the task analysis) by the sum of occurrences and nonoccurrences. The quotient was then multiplied by 100 to yield a percentage (Cooper et al., 2007).

For one participant, due to participant characteristics that cannot be disclosed for confidentiality reasons, the task analysis for the MSWO was modified to contain 23 total steps, instead of 24 total steps. Step 5, which was to ensure that the confederate was attending to the items (looking at the participant or the items), was eliminated from the task analysis and the participant was not required to engage in this step. All other 23 steps remained the same.
Interobserver Agreement

A second graduate student served as a second observer and measured participant procedural fidelity for at least 30% of sessions for all baseline and intervention conditions across all participants. The graduate student was trained by the primary researcher on the data collection process in a manner identical to that of the primary data collector. Interobserver agreement (IOA) was calculated for the dependent variable and all participants and met standards for single-case research (Kratochwill et al., 2013). When calculating IOA for participant procedural fidelity, an agreement was scored if the primary data collector and the second observer recorded the same behavior in the task analysis as a whole occurrence or a whole nonoccurrence. For example, an agreement was scored if both the primary data collector and second observer recorded “always occurred” for Step 1 on the task analysis. A disagreement was recorded if the primary data collector and the second observer did not record the same behavior in the task analysis as a whole occurrence or a whole nonoccurrence. For example, a disagreement was recorded if the primary data collector recorded “always occurred” for Step 1, but the second observer recorded “never occurred” for Step 1. IOA was calculated by dividing the total number of agreements by the sum of agreements and disagreements. The quotient was then multiplied by 100 to yield a percentage (Cooper et al., 2007).

Total IOA across all conditions was 96.7% for Riley (range: 87.5–100.0%), 99.0% for Olivia (range: 95.8–100.0%), and 99.3% for Ava (range: 95.8–100.0%). Total IOA across all conditions was 98.6% for Layla (range: 95.8–100.0%), 98.3% for Ellie (range: 95.8–100.0%), and 95.8% for Kennedy (range: 91.7–100.0%). Average instructor IOA scores for each participant, across the two conditions (i.e., baseline and intervention), are displayed in Table 2.

Primary Researcher/Confederate Procedural Fidelity

A third graduate student measured primary researcher/confederate procedural fidelity for at least 30% of sessions for all baseline and intervention conditions across all participants. The graduate student was trained by the primary researcher on the data collection process prior to the start of the study in a manner similar to that of the primary data collector. The training was identical, except that the

| Participant | Baseline (%) | Intervention (%) |
|-------------|--------------|------------------|
| Riley       | 91.7         | 100.0            |
| Olivia      | 97.9         | 100.0            |
| Ava         | 98.6         | 100.0            |
| Layla       | 97.9         | 99.0             |
| Ellie       | 97.9         | 98.6             |
| Kennedy     | 94.5         | 97.9             |
primary researcher reviewed each step of the researcher procedural fidelity data sheet with the graduate student, instead of the participant procedural fidelity data sheet.

The primary researcher/confederate procedural fidelity was the degree to which the primary researcher/confederate implemented the independent variable (i.e., email feedback), followed the session script, and engaged in the predetermined sequence behaviors (i.e., assigned confederate response data sheet) as intended (Cooper et al., 2007; Gast & Ledford, 2014). Primary researcher/confederate procedural fidelity was derived from a task analysis that depicted the behaviors the primary researcher/confederate engaged in before, during, and after each research session (see Supplementary Materials, Appendix F). Each step in the task analysis was scored “yes” if the primary researcher/confederate implemented that step correctly and “no” if the step was implemented incorrectly or was omitted. The primary researcher/confederate researcher procedural fidelity was calculated by dividing the sum of “yes” scores by the sum of “yes” plus “no” scores. The quotient was then multiplied by 100 to yield a percentage (Cooper et al., 2007). Procedural fidelity across all conditions was 100% for all participants.

**Experimental Design**

A multiple probe design across participants design was used to evaluate the effects of the MSWO preference assessment with email feedback on the participants’ procedural fidelity (Gast et al., 2014). A multiple probe design systematically introduces the independent variable (i.e., email feedback) to evaluate its effects on the dependent variable (i.e., procedural fidelity) and controls for threats to internal validity. The independent variable was introduced on one occasion for each participant, for a total of six opportunities to demonstrate experimental control and treatment effect across all six participants (Cooper et al., 2007). The multiple probe design consisted of two experimental conditions: (a) baseline and (b) intervention. Participants moved from the baseline condition to the intervention condition once visual analysis of data suggested a steady state of responding had been achieved (see Sidman, 1960). The intervention condition for each participant ended once at least five research sessions were completed and a steady state of responding had been achieved.

Additionally, based on when participants were enrolled in the study, the multiple probe design was conducted either nonconcurrent or concurrently across participants. Participants were assigned to either a non-concurrent or concurrent design based on the order in which they enrolled in the study. For the first set of participants to enroll in the study, Riley, Olivia, and Ava (see Fig. 1), the multiple probe design was nonconcurrent as they began the study on different days (i.e., Olivia started two days after Riley and Ava started six days after Olivia). For the second set of participants to enroll in the study, Layla, Ellie, and Kennedy (see Fig. 2), the multiple probe design was concurrent as they all began the study on the same day (see Slocum et al., 2022 for a primer on concurrent and nonconcurrent multiple baseline designs and variations).
Procedure

Each research session lasted approximately 15–20 min. Baseline sessions were conducted two-to-three times per week as close in days as possible, with the exception of baseline probes which consisted of at least five days (based on recommendations from Gast & Ledford, 2014) between the previous baseline session and previous baseline probe session. Identical to baseline sessions, intervention sessions were conducted two-to-three times per week as close in days as possible. All research sessions occurred at times and dates convenient for the participant and when the primary researcher was available. Research sessions always
began when the primary researcher gave the instruction to begin (i.e., “Okay, now you can begin implementing the assessment. Once you are finished, please let me know”).

All research sessions continued until: (a) the participant indicated they completed the assessment or (b) the participant did not engage in a target response from the task analysis for two min. At the end of each session, the participant was asked to display their data sheet on the screen and the primary researcher thanked the participant for attending. All research sessions were conducted via Zoom and were recorded for data collection purposes. The recording started

---

**Fig. 2** Percentage of correct implementation for three participants (Layla, Ellie, and Kennedy) across conditions.
when the participant joined Zoom and concluded immediately after the participant displayed their data sheet on the screen and the primary researcher thanked them for attending.

**Initial Research Meeting and Training**

Prior to the start of the first baseline session, all participants attended an initial research meeting with the primary researcher. The initial meeting was approximately 40 min in length. During the initial research meeting, the primary researcher reviewed the informed consent, shared a demographic questionnaire with the participant, displayed the MSWO training video in its entirety, gathered information for data sheet preference (i.e., mailed or emailed), and obtained the participant’s availability for session scheduling purposes.

**Baseline**

The purpose of this condition was to measure participant behavior prior to the introduction of the email performance-based feedback. At the beginning of each research session, participants were given up to 15 min to review the MSWO excerpt derived from DeLeon and Iwata (1996) described above. After 15 min, or when the participant reported they finished reviewing the excerpt, the participant was asked to implement the MSWO preference assessment with the primary researcher/confederate.

During implementation of the MSWO preference assessment, the primary researcher/confederate engaged in the sequence of responses determined by the assigned confederate response data sheet (see Supplementary Materials, Appendix D for an example). However, if the participant did not engage in the steps following a specific type of selection correctly, the primary researcher/confederate systematically adjusted their next selection response. For example, following a two-item sequential selection response, the participant should have engaged in Steps 16–21 on the task analysis. However, occasionally instead of ignoring the second selected item and continuing to engage with the first selected item for 30 s (Step 16), participants engaged with both of the selected items and then re-presented the remaining non-selected items. When this occurred, the primary researcher/confederate skipped the next predetermined trial (e.g., Trial #2) and moved onto the next trial (e.g., Trial #3) to match the number of items that were represented in the array. As another example, following a no selection response, if the participant continued to represent the remaining items, instead of ending that round of trials (i.e., one round equals five trials, one trial for each of the five items) and recording them as not selected (Step #22), the primary researcher/confederate continued to engage in no selection responses until the next round of trials began. Finally, as a third example, if the participant engaged in errors such as those noted above and continued to represent items in the array and the primary researcher/confederate had completed all of the predetermined trials, the primary researcher/confederate engaged in no selection responses until the participant indicated that had completed the MSWO.

After the participant indicated they had completed the MSWO (e.g., stating “I’m done”), the primary researcher/confederate provided the participant time
to calculate their results. Once the participant reported they had completed their results, the primary researcher/confederate instructed the participant to display their data sheet on the screen for a few seconds. No further instructions were given, and no feedback was provided. If a participant asked the primary researcher/confederate researcher a question during the research session, they were informed that the primary researcher/confederate could not answer their question at that time and that they should do the best they could.

After each baseline research session ended, the primary researcher sent the participant an email without any feedback components. The components that were included in the baseline emails were: (a) a general positive opening statement, (b) a request for a response (i.e., posing a scheduling question and asking for a reply), and (c) a positive closing statement. Emails were sent to participants during the baseline condition to ensure that participants were accessing, reading, and responding to emails and to further isolate the effects of feedback during the subsequent condition (see Supplementary Materials, Appendix G, for an example e-mail).

### Intervention

The purpose of this condition was to evaluate participant behavior when the participant was provided with email feedback. Experimental procedures in the intervention condition were identical to that of the baseline condition, with the exception of the email sent to participants following each research session. Email followed a format similar to that of baseline but also included components specific to supportive (i.e., the number of steps the participant engaged in correctly and comments about the participants implementation) and corrective feedback (i.e., the steps the participant did not engage in correctly). The components included in the intervention emails were: (a) a general positive opening statement, (b) supportive feedback, (c) corrective feedback, (d) a request for a response (i.e., posing a scheduling question and asking for a reply), and (e) a positive closing statement. (See Supplementary Materials, Appendix H for an example intervention email.)

### Data Analysis

Following each research session, data for each participant were graphed and reviewed by the primary researcher for trend, level, and variability of data to evaluate intervention effects (Cooper et al., 2007). Following the completion of the study, Tau-$U$ was calculated to supplement visual analysis of data and to provide a secondary measure of treatment effect. Tau-$U$ is a statistical analysis that combines non-overlap analysis between phases with trend from within the intervention phase. Tau-$U$ is dynamic in that it can calculate trend only, non-overlap between phases only, as well as combinations of trend and overlap between multiple phases (e.g., baseline and intervention conditions; Parker et al., 2011).
Results

Participant Procedural Fidelity

Nonconcurrent Set of Participants

Riley Riley’s percentage of correct implementation data are displayed in Fig. 1, top panel. In the baseline condition, Riley’s percentage of correct implementation ranged from 12.5 to 33.0% (M = 27.3%), with the last three sessions remaining stable at 33.0%. During the first intervention session, Riley’s percentage of correct implementation increased to 66.7% and then varied from 62.5 to 91.7%, until it remained stable at 100% for two sessions and 95.8% in the last session (M = 86.5%; range: 62.5–100.0%). See Supplementary Materials for raw data for Riley and the remaining participants.

Olivia Olivia’s percentage of correct implementation data are displayed in Fig. 1, middle panel. In the baseline condition, Olivia’s percentage of correct implementation was at 54.0% in the first session and then decreased and remained stable at 41.7% for the last four sessions (M = 43.8%). During the first intervention session, Olivia’s percentage of correct implementation increased to 87.5% and then remained stable between 95.8 and 100% for the last four sessions (M = 96.7%; range: 87.5–100.0%).

Ava Ava’s percentage of correct implementation data are displayed in Fig. 1, bottom panel. In the baseline condition, Ava’s percentage of correct implementation ranged from 41.7 to 54.2% (M = 48.8%). During the first intervention session, Ava’s percentage of correct implementation increased to 70.8% and then varied from 70.8 to 87.5%, until it remained stable between 95.8 and 100.0% for the last three sessions (M = 86.9%; range: 70.8–100.0%).

Concurrent Set of Participants

Layla Layla’s percentage of correct implementation data are displayed in Fig. 2, top panel. In the baseline condition, Layla’s percentage of correct implementation ranged from 0.0 to 12.5% (M = 3.34%). During the first intervention session, Layla’s percentage of correct implementation increased to 33.3% and then varied from 16.6 to 79.2%, until it remained stable between 87.5 and 95.8% for the last three sessions (M = 69.3%; range: 16.6–95.8%).

Ellie Ellie’s percentage of correct implementation data are displayed in Fig. 2, middle panel. In the baseline condition, Ellie’s percentage of correct implementation began at 30.4% in the first session and then varied between 17.4 and 21.7% across three sessions, with the last two sessions at 13.0% (M = 19.5%). During the first intervention session, Ellie’s percentage of correct implementation increased to 86.9% and then varied from 86.9 to 95.6%, until it remained stable between 95.6 and 100.0% for the last four sessions (M = 94.5%; range: 86.9–100.0%).
Kennedy Kennedy’s percentage of correct implementation data are displayed in Fig. 2, bottom panel. In the baseline condition, Kennedy’s percentage of correct implementation ranged from 25.0 to 41.6% ($M = 32.7\%)$. During the first intervention session, Kennedy’s percentage of correct implementation increased to 75.0\% and then varied from 79.2 to 91.6\%, until it remained stable at 95.8\% for the last three sessions ($M = 88.9\%;$ range: 75.0–95.8\%).

**Tau-U**

The researchers calculated Tau-U using a web-based Tau-U calculator (http://singlecase.research.org/calculators/tau-u) for the participants. The Tau-U for Riley was 1 ($p < 0.05$, $z = 2.93$), the Tau-U for Olivia was 1 ($p < 0.05$, $z = 2.74$), the Tau-U for Ava was 1 ($p < 0.05$, $z = 3.13$), the Tau-U for Layla was 1 ($p < 0.05$, $z = 3.12$), the Tau-U for Ellie was 1 ($p < 0.05$, $z = 3.10$), and the Tau-U for Kennedy was also 1 ($p < 0.05$, $z = 3.00$). Based on the weighted Tau-U for all participants, the intervention had a large or strong effect with 100\% of participants’ data showing significant improvement ($p < 0.001$, $z = 7.33$) from baseline to intervention with 95\% CIs [0.73, 1]. See Supplementary Materials for raw outputs of Tau-U calculations.

**Discussion**

Overall, the findings indicate the email performance-based feedback was effective in increasing procedural fidelity of MSWO preference assessment implementation in the pre-service teacher participants. All six participants implemented the MSWO preference assessment with high levels of procedural fidelity following email performance-based feedback. These results extend previous findings (i.e., Artman-Meecker & Hemmeter, 2013; Gomez et al., 2021; Gorton et al., 2021; Hemmeter et al., 2011; Martinez Cueto et al., 2021) by isolating the effects of email feedback. Additionally, these results support previous findings (i.e., Barton et al., 2016, 2020; Coogle et al., 2020) suggesting that email performance-based feedback alone is effective in increasing target behavior(s).

In the present study, we used a conservative “all or nothing” whole occurrence measure (similar approach used in Sipila-Thomas et al., 2021) that was biased toward deflating participant performance because a single error resulted in the entire step being marked as incorrectly implemented. An alternative measure would have been a per opportunity measure; however, the use of this measure can be biased toward inflating participant performance (Ledford et al., 2014). For example, in this study there were 15 total trials and each trial consisted of 24 different behaviors the participant could engage in. Therefore, the participant would have 24 opportunities that would be included in the calculation of a whole occurrence measure and 360 opportunities that would be included in the calculation of a per opportunity measure. When calculating the whole occurrence measure, if the participant incorrectly engaged in one specific behavior one time, during the 15 trials, and incorrectly engages in four other specific behaviors one time each, all five behaviors would
be marked incorrect and would result in a score of 79.2% or 19 out of 24 correct, whereas, if a per opportunity measure was used for each trial, the participant would have scored a 98.6% or 355 out of 360 correct. Future research may evaluate the extent to which the whole occurrence measure deflates participant performance and determine if practitioners should proceed with caution in using the whole occurrence measure over the per opportunity measure in practice.

Although the email performance-based feedback increased procedural fidelity for all six participants and Tau-U results indicated the intervention had a large or strong effect, the accuracy of responding varied across participants. Layla and Kennedy did not reach 100% procedural fidelity, though they both reached 95.8% which was implementing 23 out of the 24 steps correctly. Additionally, the number of sessions to reach high levels of fidelity varied across participants. Olivia, Ava, Ellie, and Riley reached 100% procedural fidelity after being exposed to two to six intervention research sessions, while Kennedy and Layla were exposed to six and 11 intervention research sessions, respectively, but both did not reach 100% procedural fidelity.

Future research may evaluate the extent to which the whole occurrence measure deflates participant performance and determine if practitioners should proceed with caution in using the whole occurrence measure over the per opportunity measure in practice.

When conducting a post hoc analysis of participant errors, the two steps most frequently implemented incorrectly were (1) calculating stimulus rankings (Step 23) and (2) accurately transcribing data (Step 24). Incorrectly implementing these two steps when conducting an MSWO preference assessment can be a major issue for two reasons. First, if the implementer used an item that was miscalculated as a top item when providing treatment to an individual with ASD, there may be noticeably different effects in treatment delivery because for treatment delivery to be successful (i.e., increase desired behavior or decrease undesired behavior), an effective reinforcer is required (Bottini & Gillis, 2021; Cooper et al., 2020). Second, if
the implementer did not accurately record data during the implementation of the MSWO, the inaccurate data would result in incorrect decision-making (see Cox & Brodhead, 2021) and may impact the effects of treatment delivery. However, it is unclear why these two specific steps were the two most common errors. One potential cause for the errors could be that the feedback was less effective in explaining how to correctly implement these steps. A second potential cause could be session fatigue as these two steps came at the end of the research session. Future research could evaluate if participants experience similar difficulty when calculating results and recording data when implementing other preference assessment or behavior analytic procedures. For example, a study could evaluate the use of email performance-based feedback on the implementation of a multiple stimulus with replacement preference assessment (DeLeon & Iwata, 1996), paired stimulus preference assessment (Fisher et al., 1992), or free operant preference assessment (Roane et al., 1998) and assess if participants engage in incorrect responses when calculating stimulus rankings or recording data. Additionally, future research could evaluate if another form of feedback (e.g., video model, see DiGennaro-Reed et al., 2010) is required in order to correctly calculate results.

Extension of Previous Literature

The present study extends previous research in at least three ways. First, we standardized the email responses provided to each of the participants to control for variations in the email feedback. In previous studies, emails were not standardized beyond their general frameworks (e.g., general positive opening statements, supportive feedback, corrective feedback, request for response; Artman-Meecker & Hemmeter, 2013; Barton et al., 2016, 2020; Coogle et al., 2020; Gorton et al., 2021; Hemmeter et al., 2011; Martinez Cueto et al., 2021; Zhu et al., 2021). As a result, it was unclear how researchers composed the email feedback for each participant and whether the feedback was specific or general feedback, making replication of these studies difficult. In the present study, we used a pool of specific responses for each component of the email that the primary researcher pulled from. As a result, we were able to rule out researcher variation as a potential source of influence in the present study and demonstrated that providing standardized email responses can improve target behaviors. However, comparisons between customized and standardized email performance-based feedback would still need to be conducted. Additionally, future researchers should consider evaluating other feedback components (e.g., specific vs. general feedback, frequent vs. infrequent feedback, group vs. individual feedback, and source of feedback; Novak et al., 2019; Sleiman et al., 2020) within email feedback in order to further refine email performance-based feedback interventions.

Second, we implemented the training component (i.e., the MSWO training video) prior to the baseline condition in order to isolate the effects of the email feedback and to ensure that any changes that occurred between the baseline and intervention conditions were likely a result of the email feedback alone. The results of the present study demonstrated that providing email performance-based feedback alone can improve implementation of an MSWO preference assessment. Future research could
evaluate if email feedback would increase individuals’ procedural fidelity when implementing procedures such as match-to-sample, manding, and imitation, and other behavioral interventions. Additionally, future researchers could continue to evaluate the use of email feedback with behavior analytic procedures that consist of varying number of steps of increasing difficulty to understand the conditions under which it may or may not have functional value in improving employee performance (e.g., functional analyses, functional communication training).

Third, we used a confederate with a specific list of responses to perform during each research session. Previous literature (i.e., Artman-Meecker & Hemmeter, 2013; Barton et al., 2016, 2020; Coogle et al., 2020; Gomez et al., 2021; Gorton et al., 2021; Hemmeter et al., 2011; Martinez Cueto et al., 2021) did not control for variation in the instructional environment as those studies involved classroom settings where teachers engaged with students. As a result, opportunities for participants to respond may have been at least partially affected by student behavior within those settings. By using a confederate, we eliminated or reduced variation in the instructional environment by holding errors and correct responses constant across all research sessions. Additionally, participants were exposed to each confederate response during each research session, which mimicked the responses a child with ASD may engage in during an MSWO preference assessment. As a result, we were able to evaluate participant behavior in the presence of responses likely encountered in an applied setting across the entire experiment. However, in the present study we did not assess if high procedural fidelity in the presence of a confederate would generalize to a child with ASD. Future research could evaluate if the participant’s fidelity of implementation would generalize to implementing an MSWO preference assessment with a child with ASD after receiving email feedback based on their performance implementing the MSWO with a confederate.

Limitations

Several limitations of the present study should be noted. The first limitation of this study was that it was unclear what specific components (e.g., supportive feedback, corrective feedback) of the email performance-based feedback were responsible for increases in participant procedural fidelity. Second, though all participants achieved high levels of procedural fidelity, the extent to which these gains maintain or persist long term is unknown because we only evaluated the immediate effects the email performance-based feedback had on implementation of the MSWO preference assessment. Third, the primary data collector reviewed and scored research session videos in the order in which they were recorded; as a result, the primary data collector was aware of the experimental conditions the participants were exposed to during each research session. Finally, we evaluated participant responding in the presence of a confederate instead of an individual with ASD. Although the use of a confederate allowed the researchers to control for variation in the instructional environment and provided opportunities for participants to respond to multiple learner responses, the extent to which high levels of procedural fidelity would generalize to individuals with ASD is unclear. Results of previous training studies (e.g.,
Lipschultz et al., 2015) that have used a confederate and then evaluated participant performance in the presence of a child found high levels could be achieved. However, future research could evaluate the effects of email performance-based feedback on the implementation of an MSWO preference assessment with individuals with ASD in person or within a remote context. Finally, to our knowledge, this was the first study to evaluate email performance-based feedback effects on a procedure that consists of more than eight discrete steps. Future research could continue to evaluate email performance-based feedback on other multiple step behavior analytic procedures (e.g., implementing a gross motor imitation program, conducting a functional analysis) in order to understand the extent to which email feedback can be used. Though email performance-based feedback appears to be helpful, until future research is conducted, we urge caution in viewing it as substitute for in-person feedback. Instead, email may be considered another tool to delivering high-quality feedback to help teachers improve fidelity of implementation of behavioral procedures.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10864-022-09496-z.

Funding Michigan State University’s College of Education provided financial support for participant recruitment.

Data Availability The data that support the findings of this study are available as supplementary materials.

Declarations

Conflict of interest The authors do not have any conflict of interest to disclose.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed Consent All participants provided written informed consent prior to commencement of the study.

References

Artman-Meeker, K. M., & Hemmeter, L. M. (2013). Effects of training and feedback on teachers’ use of classroom preventative practices. Topics in Early Childhood Special Education, 33(2), 112–123. https://doi.org/10.1177/02711214122447115

Barton, E. E., Fuller, E. A., & Schnitz, A. (2016). The use of email to coach preservice early childhood teachers. Topics in Early Childhood Special Education, 36(2), 78–90. https://doi.org/10.1177/0271121415244728

Barton, E. E., Velez, M., Pokorski, E. A., & Domingo, M. (2020). The effects of email performance based feedback delivered to teaching teams: A systematic replication. Journal of Early Intervention, 42(2), 143–162. https://doi.org/10.1177/1053815119872451

Barton, E. E., & Wolery, M. (2007). Evaluation of e-mail feedback on the verbal behaviors of pre-service teachers. Journal of Early Intervention, 30(1), 55–72. https://doi.org/10.1177/105381510703000105

Bottini, S., & Gillis, J. (2021). A comparison of the feedback sandwich, constructive-positive feedback, and within session feedback for training preference assessment implementation. Journal of Organizational Behavior Management, 41(1), 83–93. https://doi.org/10.1080/01608061.2020.1862019

Brodhead, M. T. (2022). Training video. osf.io/xhdzw
Chazin, K. T., & Ledford, J. R. (2016). Multiple stimulus without replacement (MSWO) preference assessment. In Evidence-based instructional practices for young children with autism and other disabilities. Retrieved from http://ebip.vkcisites.org/multiple-stimulus-without-replacement

Coogle, C. G., Ottley, J. R., Storie, S., Rahn, N. L., & Kurowski-Burt, A. (2017). eCoaching to enhance special educator practice and child outcomes. Infants and Young Children, 30(1), 58–75. https://doi.org/10.1097/YIC.0000000000000682

Coogle, C. G., Ottley, J. R., Storie, S., Rahn, N. L., & Kurowski-Burt, A. (2020). Performance-based feedback to enhance preservice teachers’ practice and preschool children’s expresssive communication. Journal of Teacher Education, 72(2), 188–202. https://doi.org/10.1177/0022487118803583

Coogle, C. G., Rahn, N. L., Ottley, J. R., & Storie, S. (2016). eCoaching across routines to enhance teachers’ use of modeling. Teacher Education and Special Education, 39(4), 227–245. https://doi.org/10.1177/088406415621959

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). Applied behavior analysis (2nd ed.). Pearson Education.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). Applied behavior analysis (3rd ed.). Pearson Education.

Cox, D. J., & Brodhead, M. T. (2021). A proof of concept analysis of decision-making with time-series data. The Psychological Record, 71(2), 349–366. https://doi.org/10.1007/s00732-020-00451-w

DeLeon, I. G., & Ivata, B. A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. Journal of Applied Behavior Analysis, 29(4), 519–533. https://doi.org/10.1901/jaba.1996.29.519

DeLeon, I. G., Ivata, B. A., & Roscoe, E. M. (1997). Displacement of leisure reinforcers by food during preference assessments. Journal of Applied Behavior Analysis, 30(3), 475–484. https://doi.org/10.1901/jaba.1997.30-475

DiGennero-Reed, F. D., Coddin, R., Catania, C. N., & Maguire, H. (2010). Effects of video modeling on treatment integrity of behavioral interventions. Journal of Applied Behavior Analysis, 43(2), 291–295. https://doi.org/10.1901/jaba.2010.43-291

Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. Journal of Applied Behavior Analysis, 25(2), 491–498. https://doi.org/10.1901/jaba.1992.25-491

Gast, D. L., & Ledford, J. R. (2014). Single case research methodology: Applications in special education and behavioral sciences (2nd ed.). Routledge.

Gast, D. L., Lloyed, P. B., & Ledford, J. R. (2014). Multiple baseline and multiple probe designs. In D. L. Gast & J. R. Ledford (Eds.), Single subject research methodology: Applications in special education and behavioral sciences (pp. 252–296). Routledge.

Gomez, L., Barton, E. E., Winchester, C., & Locchetta, B. (2021). Effects of email performance feedback on teachers’ use of play expansions. Journal of Early Intervention, 43(3), 235–254. https://doi.org/10.1177/1053815120969821

Gorton, K., Alladay, R. A., Lane, J. D., & Ault, M. J. (2021). Effects of brief training plus electronic feedback on increasing quantity and intonation of behavior specific praise among preschool teachers. Journal of Behavioral Education. https://doi.org/10.1007/s10864-020-09427-w

Graff, R. B., & Karsten, A. M. (2012). Assessing preferences of individuals with developmental disabilities: A survey of current practices. Behavior Analysis in Practice, 5(2), 37–48. https://doi.org/10.1007/bf03391822

Groskreutz, N. C., Groskreutz, M. P., & Higbee, T. S. (2011). Effects of varied levels of treatment integrity on appropriate toy manipulation in children with autism. Research in Autism Spectrum Disorders, 5(4), 1358–1369. https://doi.org/10.1016/j.rasd.2011.01.018

Hemmeter, M. L., Snyder, P., Kinder, K., & Artman, K. (2011). Impact of performance feedback delivered via electronic mail on preschool teachers’ use of descriptive praise. Early Childhood Research Quarterly, 26(1), 96–109. https://doi.org/10.1016/j.ecresq.2010.05.004

Iwata, B. A., Wallace, M. D., Kahng, S., Lindberg, J. S., Roscoe, E. M., Conners, J., Hanley, G. P., Thompson, R. H., & Worsdell, A. S. (2000). Skill acquisition in the implementation of functional analysis methodology. Journal of Applied Behavior Analysis, 33(2), 181–194. https://doi.org/10.1901/jaba.2000.33-181

Kang, S., O’Reilly, M., Lancioni, G., Falcomata, T. S., Sigafos, J., & Xu, Z. (2013). Comparison of the predictive validity and consistency among preference assessment procedures: A review of the literature. Research in Developmental Disabilities, 34(4), 1125–1133. https://doi.org/10.1016/j.ridd.2012.12.021
Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. Remedial and Special Education, 34, 26–38. https://doi.org/10.1177/0741932512452794

Krick Oborn, K. M., & Johnson, L. D. (2015). Coaching via electronic performance feedback to support home visitors’ use of caregiver coaching strategies. Topics in Early Childhood Special Education, 35(3), 157–169. https://doi.org/10.1177/0271124115592411

Leaf, J. B., Leaf, R., Alcalay, A., Leaf, J. A., Ravid, D., Dale, S., Kassardjian, A., Tsuji, K., Taubman, M., McEachin, H., & Oppenheim-Leaf, M. (2015). Utility of formal preference assessment for individuals diagnosed with autism spectrum disorder. Education and Training in Autism and Developmental Disabilities, 50(2), 199–212.

Ledford, J. R., Wolery, M., & Gast, D. L. (2014). Controversial and critical issues in single case research. In D. L. Gast & J. R. Ledford (Eds.), Single subject research methodology: Applications in special education and behavioral sciences (pp. 377–396). Routledge.

Lill, J. D., Shriver, M. D., & Allen, K. D. (2021). Stimulus preference assessment decision-making system (SPADS): A decision-making model for practitioners. Behavior Analysis in Practice, 14, 1144–1156. https://doi.org/10.1007/s40617-020-00539-3

Lipschultz, J. L., Vladescu, J. C., Reeve, K. F., Reeve, S. A., & Dipsey, C. R. (2015). Using video modeling with voiceover instruction to train staff to conduct stimulus preference assessments. Journal of Developmental and Physical Disabilities, 27(4), 505–532. https://doi.org/10.1007/s10882-015-9434-4

Maple Tech International LLC. (2022). Random number generator. https://www.calculator.net/random-number-generator.html

Martinez Cueto, A. P., Barton, E. E., & Bancroft, J. C. (2021). The effects of training and performance feedback on preservice teachers’ use of statements that promote preschool children’s social interactions. Journal of Positive Behavior Interventions. https://doi.org/10.1177/1098300721994200

Miltenberger, R. G. (2012). Behavior skills training procedures. In R. G. Miltenberger (Ed.), Behavior modification: Principles and procedures (5th ed., p. 217). Wadsworth.

Novak, M. D., DiGennaro Reed, F. D., Erath, T. G., Blackman, A. L., Ruby, S. A., & Pellegrino, A. J. (2019). Evidence-based performance management: Applying behavioral science to support practitioners. Perspectives on Behavior Science, 42(4), 955–972. https://doi.org/10.1007/s40614-019-00232-z

O’Handley, R. D., Pearson, S., Taylor, C., & Congdon, M. (2021). Training preservice school psychologists to conduct a stimulus preference assessment. Behavior Analysis in Practice, 14(2), 445–450. https://doi.org/10.1007/s40617-020-00537-5

Park, J., Johnson, D. A., Moon, K., & Lee, J. (2019). The interaction effects of frequency and specificity of feedback on work performance. Journal of Organizational Behavior Management, 39(3–4), 164–178. https://doi.org/10.1080/01608061.2019.1632242

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single case research: Tau-U. Behavior Therapy, 42(2), 284–299. https://doi.org/10.1016/j.beth.2010.08.006

Roane, H. S., Vollmer, T. R., Ringdahl, J. E., & Marcus, B. A. (1998). Evaluation of a brief stimulus preference assessment. Journal of Applied Behavior Analysis, 31(4), 605–620. https://doi.org/10.1901/jaba.1998.31-605

Rosales, R., Gongola, L., & Homlitas, C. (2015). An evaluation of video modeling with embedded instructions to teach implementation of stimulus preference assessments. Journal of Applied Behavior Analysis, 48(1), 209–214. https://doi.org/10.1002/jaba.174

Roscoe, E. M., Fisher, W. W., Glover, A. C., & Volkert, V. M. (2006). Evaluating the relative effects of feedback and contingent money for staff training of stimulus preference assessments. Journal of Applied Behavior Analysis, 39(1), 63–77. https://doi.org/10.1901/jaba.2006.7-05

Schles, R. A., & Robertson, R. E. (2019). The role of performance feedback and implementation of evidence-based practices for preservice special education teachers and student outcomes: A review of the literature. Teacher Education and Special Education, 42(1), 36–48. https://doi.org/10.1177/0888406417736571

Shapiro, M., & Kazemi, E. (2017). A review of training strategies to teach individuals implementation of behavioral interventions. Journal of Organizational Management, 37(1), 32–62. https://doi.org/10.1080/01608061.2016.1267066

Sidman, M. (1960). Tactics of scientific research: evaluating experimental data in psychology. Authors Cooperative.
Sipila-Thomas, E. S., Brodhead, M. T., Plavnick, J. B., & White, A. N. (2021). An evaluation of fidelity of implementation of a manualized social-play intervention. *Journal of Developmental and Physical Disabilities, 33*, 391–412. https://doi.org/10.1007/s10882-020-09754-w

Sleiman, A. A., Sigurjonsdottir, S., Elnes, A., Gage, N. A., & Gravina, N. E. (2020). A quantitative review of performance feedback in organizational settings (1998–2018). *Journal of Organization Behavior Management, 40*(3–4), 303–332. https://doi.org/10.1080/01608061.2020.1823300

Slocum, T. A., Pinkelman, S. E., Joslyn, P. R., & Nichols, B. (2022). Threats to internal validity in multiple-baseline design variations. *Perspectives on Behavior Science*. https://doi.org/10.1007/s40614-022-00326-1

Vladescu, J. C., Caroll, R., Paden, A., & Kodak, T. M. (2012). The effects of video modeling with voiceover instruction on accurate implementation of discrete-trial instruction. *Journal of Applied Behavior Analysis, 45*(2), 419–423. https://doi.org/10.1901/jaba.2012.45-419

Warrilow, G. D., Johnson, D. A., & Eagle, L. M. (2020). The effects of feedback modality on performance. *Journal of Organizational Behavior Management, 40*(3–4), 233–248. https://doi.org/10.1080/01608061.2020.1784351

Zhu, J., Bruhn, A., Yuan, C., & Wang, L. (2021). Comparing the effects of videoconference and email feedback on treatment integrity. *Journal of Applied Behavior Analysis, 54*(2), 618–635. https://doi.org/10.1002/jaba.810

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.