Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Comparison of In-Person and Virtual Behavior Analytic Supervision in a Preschool Setting

Robin Nuzzolo

robinnuzzolo@yahoo.com

Fred S. Keller School

Du Lin

Fred S. Keller School

Buttigieg Susan

Fred S. Keller School

Katharine Loomis

Fred S. Keller School

Faheema Abdool-Ghany

National University

R. Douglas Greer

Fred S. Keller School

Research Article

Keywords: Learn unit, Teacher Performance Rate and Accuracy, teacher training, virtual instruction

Posted Date: August 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10553118/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

This study compared the effectiveness of virtual and in-person teacher training delivered by behavior analyst supervisors on students' curricular goals. The participants were four newly hired, female teacher assistants in a preschool setting. During baseline, experimenters provided a model and demonstration of learn units at the start of each session without offering feedback. During intervention, experimenters delivered contingent feedback either via Zoom or in person using the Teacher Performance Rate and Accuracy (TPRA) measure. A post-intervention probe was conducted with a different child after the participant demonstrated mastery during intervention. A counterbalanced multiple-treatment design across participants was used to evaluate the effects of the two training conditions. Results showed that both virtual and in-person training were equally effective in improving the participants' accurate learn unit presentation.

Introduction

In 2020, the Behavior Analyst Certification Board (BACB) introduced guidelines for virtual behavior analytic supervision, in response to the COVID-19 pandemic (Simmons et al., 2021). Implementing applied behavior analytic (ABA) supervision virtually, utilizing telehealth, has many advantages for both the supervisor and the trainee. Telehealth is defined as "the use of electronic information and telecommunication technologies to support long-distance clinical health care, patient and professional health-related education, public health and health administration" (American Telemedicine Association, 2017). It offers options for supervision in rural areas which may not have otherwise been feasible for some individuals prior to telehealth (Britton & Cicora, 2019; Simmons et al., 2021; Wood et al. 2005). The cost for ABA supervision via telehealth is most often lower than that of in-person supervision services (Figge et al., 2022; Hartzler et al., 2022; Ivanova et al., 2025; Lindgren et al., 2016), also allowing supervisors more flexible schedules and availability when cutting down on traveling time to and from various work sites (Perle & Zheng, 2023; Sullivan et al., 2022).

Although virtual supervision offers advantages, it does not come without challenges for all involved. One key benefit, as noted earlier, is its ability to expand access to supervision and services in rural and underserved areas. However, this advantage may be offset by barriers such as limited internet services or reliable wireless networking technology (Wi-Fi) (Chen et al., 2020; Sipila-Thomas & Brodhead, 2023). Additional difficulties can arise when supervising a mentee virtually that would otherwise be evident when in person. For example, it may be difficult to view all materials used on a tabletop or for physical prompting procedures (Simmons et al., 2021). Similarly, if the mentees or their clients have low volume voices it may be more difficult for a supervisor to hear what is being said, especially when with other background noise in the room (Lerman et al., 2020). In addition, some suggest that virtual supervision can create a less personal experience as opposed to face-to-face interactions (Wilczenski & Coomey, 2005). The literature regarding best practices for providing behavior analytic supervision via telehealth has increased (Sipila-Thomas & Brodhead, 2023) which may provide guidance to help mitigate many of the barriers.

In a survey of 69 board certified behavior analysts (BCBA) and board-certified behavior analysts with a doctoral designation (BCBA-D), Sipila-Thomas and Brodhead (2023) identified two of the most frequently reported barriers to telehealth supervision: unreliable internet connectivity and difficulty modeling or demonstrating strategies in a manner that supervisees could clearly observe and understand. The former may interfere with the supervisee participating fully in the training, and the latter may result in poor instructor fidelity or treatment outcomes for the student. Based on these findings, the authors called for empirical research for telehealth supervision, with specific emphasis on identifying effective strategies to address these problems brought to light in their survey.

Robertson et al. (2023) used telehealth to train three behavior technicians (BTs) who provided one-on-one intervention to children with autism spectrum disorder (ASD) at an agency located approximately 80 km away through videoconferencing software. The dependent variables were the treatment fidelity of an errorless learning procedure and a preference assessment. The independent variable was a behavior skills training (BST) protocol delivered via a telecommunication platform. During training, the supervisor provided real-time, vocal feedback to the BTs through a bug-in-ear device. Baseline data showed the participants, despite having worked at the agency for six months prior, correctly implemented the procedures on only 40%–70% of opportunities. Following telehealth BST training, their performance increased substantially to 70%–100% accuracy.

Akers et al. (2024) also demonstrated the effectiveness of telehealth for training seven graduate students to implement caregiver coaching with discrete trial teaching (DTT). The researchers evaluated participants' accuracy in implementing caregiver coaching procedures before and after receiving remote delayed performance feedback. Following each session, the experimenter met virtually with the participant to provide feedback using praise for correct steps, corrective feedback for errors and opportunities for role-play and questions. During baseline, all seven participants implemented fewer than 50% correct steps. Following virtual supervisor feedback all participants achieved 100% correct with implementation fidelity.

To our knowledge, only one peer reviewed article has directly compared virtual and in-person behavior analytic supervision. This study evaluated one instructional component, the correction. Pantermuehl and Lechago (2015) compared in-person supervision, versus virtual with three ABA therapists providing one-to-one home-based instruction to students with ASD as part of a clinic-based program. During the in-person condition, the supervisor positioned themselves a few feet behind the participant and their student’s cubicle and provided real-time, moment to moment feedback on the behavior analyst’s DTT correction procedures. During the virtual supervision sessions, the feedback was delivered procedurally identical to that of the in-person sessions except the supervisor was visible on an iPad screen via Skype™. Results showed that for all three participants in-person and virtual supervision had almost identical positive effects on the therapists’ data in which correct DTT increased significantly.

Effective and ethical supervision is essential to those directly providing behavior analytic services. As the demand for qualified supervision to support those accruing hours and experience toward BCBA and

Registered Behavior Technicians (RBT) continues to grow, identifying effective and efficient supervision practices has become increasingly important (Andzik & Kranak, 2021). Moreover, “improper supervision” has been among the most frequently reported violation reported to the BACB (BACB, 2018). More research is needed in this area. The present study compared the effects of supervision delivered through traditional in-person style with that delivered virtually. To our knowledge, it is the first of its kind to measure all components of instruction in a school setting.

Method

Participants

Six newly hired teacher assistants participated in the study (see Table 1). All participants were female, full-time employees at a preschool that implemented instruction based on the principles of applied behavior analysis (ABA). None of the participants had prior experience teaching, utilizing behavior analytic methodology or providing instruction. Each participant was assigned to a different classroom and received supervision from a different doctoral-level behavior analyst supervisor.

Target students for each teacher assistant participant were selected based on their ability to attend to instruction with minimal interfering behaviors. The experimenters were all Senior Behavior Analyst supervisors in the school and had completed measurable advanced competence in the strategic science of teaching and its supervision (CABAS® Behavior Analyst Supervisor Ranks, www.scienceofteaching/ranks.org). They all had 10-20 years of experience in behavior analysis prior to the study, and their observation and feedback were a typical occurrence in the setting. The study was approved by the participants’ school IRB board. Informed consents were obtained from all participants prior to the study.

Setting

All participants were employed by the same private, publicly funded preschool for neurodiverse and neurotypical students, which utilized an applied behavior analytic methodology for teaching. Students ranged in age from 18 months to 5 years and attended school five days per week five hours per day. The experiment was conducted in a quiet and minimally furnished research room in the school. The research room was located close to the experimenters’ office (approximately 15 feet away), allowing the experimenter to enter immediately if needed. The room contained two child-sized tables and three chairs. An iPad mounted on a stand was used for teleconferencing and to record the sessions for interobserver agreement (IOA) purposes later.

During virtual training sessions, the participant, each novice teaching assistant, presented instructions to the child at a table in the research room while the experimenter joined remotely via Zoom using the mounted iPad. During in-person sessions, the experimenter was present in the room with the participant and the child. The participant delivered instruction to the student at a child size table. There were no other students or staff in the room at the time of the study. For Dyad 1 (Participants A & B), the

participant delivered instruction on tacting animal pictures, following two-step vocal directions, tracing lines, and imitating two-step gross motor movements. For Dyad 2 (Participants C & D), the participant delivered instruction on tracing lines, imitating two-step gross motor movements, tacting pictures, and following two-step vocal directions. For Dyad 3, the participant delivered instruction on object use imitation, pegboard manipulation, following one-step directions and pointing to pictures of animals. All of these are part of a tested criterion-based curriculum and curriculum assessment (Singer-Dudek et al., 2023).

Materials & Equipment

The experimenters used a 9-inch iPad to present video demonstrations. The iPad was mounted on a floor stand positioned in front of the participant's table to ensure clear visibility. During virtual sessions, the experimenter initiated a Zoom meeting on the iPad and recorded each session using iCloud. Prior to each session, the experimenters prepared all necessary materials, including data collection sheets, a clipboard, worksheets, pens, and markers. The participant prepared the student's preferred edible reinforcers.

For tact instruction, picture stimuli were presented on an iPad using PowerPoint slides displayed in full-screen mode. For the object use imitation program, the experimenter prepared the required toys and objects in advance. For the tracing symbols program, worksheets were created ahead of time. Each letter-sized page was divided into four quadrants by intersecting lines, with one symbol presented in each section. A total of five symbols were used, each repeated four times. The student used a marker to trace over the lines.

Dependent Variable and Data Collection

The dependent variable was the number of correctly presented learn units emitted by the participants. A correct learn unit required the participant to (a) present the correct antecedent, (b) accurately record the student's response, and (c) deliver the appropriate consequence contingent on the student's response (i.e., reinforcement following correct responses or the correction procedure following incorrect responses) (Greer & Hogan-McDonough, 1999).

Using the TPRA form, the experimenters recorded the accuracy of each component of the participants' learn units (antecedent, student response, and the reinforcement or correction operation) based on information gathered from the participant prior to the onset of the observation. The TPRA form was paired with the participants' data sheet and numbered 1-20. The experimenter collected their data in the boxes labeled "LU/trial" in the same vertical fashion the participant collected their data for the student on a data sheet numbered 1-20, adding codes for the antecedent, response and reinforcement or correction procedure. The experimenter used the following notation: ✓ for antecedents, + or – for student behavior, and R or C for the participant's response to student behavior. Any components that were incorrectly delivered or omitted were circled. If any component was incorrect, the entire learn unit was marked as incorrect. Any comments regarding that particular learn unit were recorded next to its

corresponding number under the column marked “comments”. The experimenter’s timer was stopped at the completion of the instructional program.

Design

We used a counterbalanced multiple-treatment design across participant dyads to compare the effects of virtual and in-person supervision. Counterbalancing controlled for potential order effects and ensured that each training condition was experienced equally across participants and curricular goals. Participants were arranged into three dyads: Dyad 1 (Participants A and B), Dyad 2 (Participants C and D), and Dyad 3 (Participants E and F). Table 2 presents the sequence of training conditions and curricular goals for each participant.

Each participant completed two experimental conditions. Each condition involved a different set of instructional programs. For each experimental condition, participants completed baseline before receiving intervention in either virtual or in-person supervision. Following each intervention phase, a post-intervention probe was conducted with a novel student. They then transitioned to the other experimental condition and were trained on the second instructional program.

Within each dyad, the order of supervision conditions and curricular goals was counterbalanced. In Dyad 1, Participant A first received virtual supervision on tacting pictures and following directions, whereas Participant B first received in-person supervision with the goals. Upon reaching mastery, both participants were trained on tracing lines and gross motor imitation under the alternate condition. In Dyad 2, Participant C began with virtual supervision on tracing lines and gross motor imitation, while Participant D began with in-person supervision on the same goals. Following mastery, both participants switched to tact pictures and follow directions under the alternate condition. In Dyad 3, Participant E initially received virtual supervision on object use imitation and peg manipulation, whereas Participant F received in-person supervision on the same goals. After meeting mastery criteria, both participants were trained on follow directions and point to colors under the alternate condition.

Procedures

Prior to the onset of the study, the participants observed daily classroom activities in their assigned classroom for three days. During this period, they may have observed instruction from other teachers without direct training. They also did not independently deliver instruction to students.

Procedures for teaching students the curricular goals included in the present study were as follows: In the following one-step vocal directions program, the participant delivered a single-step instruction (e.g., stomp feet, wave) and allowed 3 s for a response. A correct response was defined as the student accurately completing the instruction within 3 s. Incorrect responses included the student emitting inaccurate response or failing to emit any response within 3 s.

In the following two-step vocal directions program, the participant delivered a two-step instruction (e.g., Stand up and clap your hands) and allowed 3 s for a response. A correct response was defined as the

student completing both steps, regardless of the sequence, within 3 s. Incorrect responses included the student completing only one step, responding inaccurately, or failing to respond within 3 s.

In the tracing symbols program, the participant presented a worksheet containing four symbols per page. A correct response was defined as tracing within one-quarter inch of the line. Incorrect responses included tracing beyond one-quarter inch of the line or failing to respond within 3 s.

In the two-step gross motor imitation program, the participant delivered a vocal instruction paired with a modeled two-step motor sequence (e.g., “Do this”). A correct response occurred when the student imitated both actions, regardless of sequence, within 3 s. Incorrect responses included imitating only one step, inaccurate imitation, or no response within 3 s.

In the object use imitation program, the participant presented a toy or object, modeled a functionally relevant action, and delivered the instruction “Do this.” A correct response was defined as accurately imitating the modeled action within 3 s. Incorrect responses included emitting a different action or failing to respond within 3 s.

In the pegboard manipulation program, the participant modeled placing a specified number of pegs into a pegboard and delivered the instruction “Do this.” A correct response was defined as imitating the same action using the same number of pegs. Incorrect responses included using a different number of pegs, manipulating materials differently, or failing to respond within 3 s.

In the pointing to animal pictures program, the participant presented an array of three pictures and instructed the student to point to a specified animal. A correct response was defined as pointing to the correct picture within 3 s. Incorrect responses included pointing to an incorrect picture or failing to respond within 3 s.

In-Person Baseline

During baseline, each participant was assigned two instructional goals. At the start of each instructional session, the participant observed the supervisor accurately model demonstration of a learn unit presentation. A learn unit is a teacher student instructional trial that meets certain criteria (Albers & Greer, 1991). It begins with the teacher obtaining the student’s attention. The teacher presents an antecedent (material with or without vocal direction). If the student responds correctly within a specific time, in this case 3 s, the teacher delivers reinforcement (e.g., vocal praise, preferred edibles, preferred activity, tokens or other prosthetic reinforcers) contingent on the correct response, records the data and concludes the learn unit.

If the student failed to respond within 3 s or emitted an incorrect response, the teacher modeled the correct response and again presents the antecedent to the student and recorded the data. If the student responded correctly within 3 s during correction, the teacher confirmed the correct response with a neural tone (e.g., “That is correct.”) without providing any reinforcement. If the student continued to emit incorrect response or failed to respond, the teacher modeled the correct response again and repeated

the procedure until the student responded correctly and independently, or until three times during the correction procedure before moving on to the next learn unit.

Each model demonstration consisted of five learn units and included a combination of correct and incorrect student responses. This arrangement ensured the participants had an opportunity to observe the contingencies associated with both correct and incorrect responses. During the demonstration, participants were instructed to record data on the each observed learn unit. At the conclusion of the demonstration, the supervisor reviewed and calibrated the participants' data recording for accuracy.

Following the model demonstration, the participant conducted one independent teaching session with the same student. Each session consisted of 20 learn units, including five instructional targets, each with four learning opportunities. No feedback was provided to the participant during baseline sessions. All sessions were video recorded and uploaded to YouTube for interobserver agreement (IOA) assessment. The supervisor totaled the number of correct learn units at the end of each session. Steps 2 and 3 were repeated for a minimum of three sessions or until stable data were observed across sessions.

In-Person Training

During the in-person training phase the experimenter, an SBA and BCBA-D, sat next to the participant and observed them working with one student, and conducted a *Teacher Performance Rate and Accuracy* (TPRA) measure (Nuzzolo et al., 2025; Ross et al., 2005) on their delivery of a 20 learn unit instructional program. Across all curricular goals, the participant presented the antecedent and then provided the student with 3s to respond. When an incorrect or no response occurred, the participant implemented a correction procedure consisting of modeling the correct response followed by an additional independent opportunity. This sequence was repeated for up to three attempts.

The top section of the TPRA form was then filled out, writing in their initials, the date, the participant's initials, the curricular goal, and the learning objective. The supervisor provided participants with a data collection sheet with the current short-term objective and the necessary instructional materials as well as reinforcers for the student (e.g., preferred edibles, instructions on physical touch, tokens etc.). During the TPRA condition, the experimenter observed the participant delivering antecedents to the student, recording data, and providing feedback in the form of reinforcement or correction procedures throughout the session. They provided immediate vocal feedback to the participants when errors occurred in any component of instruction and offered specific praise for correct learn units during natural breaks, such as: "That was excellent! I liked how you paired physical contact and edibles with his listening to voices." The experimenters ensured that they provided corrective feedback for all errors recorded on the TPRA form. For example, if the participant reinforced a correct response during a correction procedure, the experimenter would clarify: "We do not provide reinforcement for incorrect responses during the correction operation." Additional guidance may have involved changing instructional procedures, modelling a correct procedure, prompting the participant to perform a procedure, or providing further explanation. After the end of the session, the experimenter provided overall feedback, and the participant received a copy of the TPRA form with specific written feedback.

Virtual Baseline

The experimenters recorded videos showing a BCBA-D working with a preschool-aged student on the selected curricular goals. Each video contained five learn units with both correct and incorrect responses. The duration of the videos was from 1.4 to 7.2 min with an average of 2.4 min. The videos were presented at the beginning of each intervention session for participants in the virtual training condition regardless of how many opportunities it took the participant to meet criterion for that phase. The supervisor then observed the participant working with a student for 20 learn units on Zoom without providing any feedback.

Virtual Training

During the virtual training phase, the experimenter created a Zoom link and shared it with the participant. The experimenter started the virtual session from her office computer and waited for the participant to join. The experimenter first presented pre-recorded video models of the target goals via Zoom. The participant viewed these videos independently in the research room. The arrangement was designed to minimize competing demands of student behavior management and allowed the participant to focus without distraction.

After viewing the video models, the participant retrieved a student from their classroom and prepared to begin the session. Prior to instruction, the experimenter positioned the iPad to ensure clear visibility of the participant, the student, and the instructional materials, as well as clear audio throughout the session. The supervisor then greeted them and observed the session as the participant delivered learn units while implementing the TPRA procedure in the same manner as described in the in-person training.

The experimenter provided immediate performance feedback as appropriate and conducted point-by-point data verification with the participant to assess interobserver agreement (IOA) for each trial. The mastery criterion for this phase was defined as achieving at least 90% accuracy across two consecutive sessions. The experimenter used the fidelity checklist to make sure they completed all steps.

Post-Intervention Probe

Once a participant met the mastery criterion, a post-intervention probe was conducted. For this probe, the experimenter selected a new student for the participant to teach using the same two curricular goals under the same condition. No feedback or consequences were provided to the participant during this probe (identical to baseline). The experimenter did not provide any model or demonstration of the learn unit presentation. The post-intervention probe was conducted to assess generalization and the transformation of stimulus control with novel students.

Treatment Fidelity and Interobserver Agreement (IOA)

Another independent observer watched the recorded sessions and recorded the accuracy of the experimenter's implementation of the procedures. The experimenters videotaped the sessions for IOA.

Another experimenter, a BCBA-D, watched the recorded videos and calibrated the data trial by trial. IOA was calculated by using the total number of agreements divided by the total number of agreements plus disagreements, multiplying by 100.

During the virtual baseline condition, IOA was conducted for 70% of Participant A's baseline sessions with 100% agreement, for 40% of Participant B's baseline sessions with 100% agreement, for 66% of Participant C's baseline sessions with a mean of 97.5% agreement (range: 95%-100%), for 33% of Participant D's baseline sessions with 100% agreement, for 33% of Participant D's baseline sessions with 100% agreement, for 50% of Participant E's baseline sessions with 100% agreement, and for 33% of Participant F's baseline sessions with 100% agreement.

During the virtual intervention condition, IOA was conducted for 100% of Participant A's sessions with 100% agreement, for 83% of Participant B's sessions with 99% agreement (range: 95%-100%), for 75% of Participant C's sessions with 100% agreement, for 50% of Participant D's sessions with 100% agreement, for 33% of Participant D's sessions with 100% agreement, and for 16% of Participant F's sessions with 100% agreement.

During the in-person baseline condition, IOA was conducted for 70% of Participant A's baseline sessions with 100% agreement, for 40% of Participant B's baseline sessions with 100% agreement, for 33% of Participant C's baseline sessions with a mean of 100% agreement, for 33% of Participant D's baseline sessions with 100% agreement, for 33% of Participant D's baseline sessions with 100% agreement, for 40% of Participant E's baseline sessions with 100% agreement, and for 33% of Participant F's baseline sessions with 100% agreement.

During the in-person intervention condition, IOA was conducted for 100% of Participant A's sessions with a mean of 96.3% agreement (range: 90%-100%), for 40% of Participant B's sessions with 100% agreement, for 33% of Participant C's sessions with 100% agreement, and for 50% of Participant E's sessions with 100% agreement.

To assess treatment fidelity, an independent observer evaluated the experimenter's implementation of the intervention to ensure adherence to the procedure. The observer scored the experimenter's implementation of procedure during baseline and intervention sessions in both conditions. Treatment fidelity was scored 100% across all observed sessions. Please see Table 5 for the percentage of sessions for treatment fidelity in each condition.

Social Validity

We assessed social validity for both participants and experimenters using different questionnaires administered following completion of the study. Each questionnaire consisted of items rated on a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree) designed to evaluate the perceived importance, acceptability, feasibility, and effectiveness of the training procedures (please see Appendix). The participant questionnaire included items related to the importance of systematic teacher training,

ease of learning both in-person and virtual training formats, usefulness of modeling demonstrations and feedback, perceived improvements in instructional accuracy, overall satisfaction, and self-reported confidence in delivering instruction. The experimenter questionnaire assessed similar measures including ease of implementation, efficiency of feedback delivery, effectiveness in improving participant performance, feasibility within classroom routines, and overall satisfaction with both training formats. Both questionnaires also included an item assessing preference for training format (in-person, virtual, hybrid, or no preference).

Results

During baseline under the virtual training condition, Participant A's correct responses ranged from 12 to 17 for tact animals, and for two-step follow-up directions, from 5 to 11 across five sessions. With the introduction of virtual training with TPRA feedback, her performance improved immediately. She met mastery for tact animals within two sessions and for two-step follow-up directions within three sessions. During the generalization probe with a novel student, she maintained high accuracy, achieving 100% for tact animals and 80% for follow-up directions.

When switched to in-person training condition, her correct responses ranged from 7 to 16 tracing marks, and from 14 to 18 for two-step gross motor imitation across sessions during baseline. Following in-person training with TPRA feedback, she reached mastery for tracing marks within two sessions and for gross motor imitation within three sessions. During the subsequent generalization probe, she maintained 100% accuracy across both curricular goals (please see Fig. 1).

During baseline under the in-person training condition, Participant B's correct responses ranged from 2 to 14 across sessions for tact animals, and for two-step follow-up directions, from 12 to 15 across five sessions. With the introduction of virtual training with TPRA feedback, her correct responses improved immediately. Participant B met mastery for tact animals within four sessions and for two-step follow-up directions within six sessions. During the generalization probe with a novel student, Participant B maintained high accuracy, achieving 100% for tact animals and 80% for follow-up directions.

In baseline under the virtual training condition, Participant B's correct responses ranged from 0 to 14 for tracing marks, and for two-step gross motor imitation, from 14 to 17 across three sessions. Following in-person training with TPRA feedback, Participant B reached mastery for tracing marks within two sessions and for gross motor imitation within three sessions. During the subsequent generalization probe, performance was maintained at 95% and 85% accuracy across curricular goals (please see Fig. 1).

During baseline under the in-person training condition, Participant C's correct responses remained low, ranged from 0 to 5 across sessions for tracing lines, and for two-step gross motor imitation, from 2 to 3 across three sessions. With the introduction of virtual training with TPRA feedback, her correct responses improved immediately. Participant C met mastery for tracing lines within four sessions and for two-step gross motor imitation within five sessions. During the generalization probe with a novel

student, Participant C maintained high accuracy, achieving 90% for tracing lines and 80% for two-step gross motor imitation.

In baseline under the virtual training condition, Participant C's correct responses ranged from 0 to 14 for tacting pictures, and for following two-step vocal directions, from 9 to 19 across six sessions. Following in-person training with TPRA feedback, Participant C reached mastery for tacting pictures within three sessions and for following two-step vocal directions within four sessions. During the subsequent generalization probe, performance was maintained at 100% and 95% accuracy across curricular goals (please see Fig. 2).

During baseline under the in-person training condition, Participant D's correct responses for tracing lines remained at 0 across three sessions; therefore, the experimenter discontinued the program. In contrast, she demonstrated improvement with in-person modeling for two-step gross motor imitation, reaching mastery within five sessions. With the introduction of virtual training with TPRA feedback, her performance on tracing lines improved immediately, and she met mastery within six sessions. During the subsequent generalization probe with a novel student, she maintained 100% accuracy across both curricular goals.

In baseline under the virtual training condition, Participant D met mastery without intervention, reaching criterion within two sessions for two-step follow-up directions and within four sessions for tacting pictures. During the subsequent generalization probe, performance was maintained at 100% accuracy across both curricular goals (see Fig. 2).

During baseline under the in-person training condition, Participant E's correct responses remained highly variable, ranged from 8 to 14 for object use imitation, and for manipulating pegs, from 6 to 14 across five sessions. With the introduction of virtual training with TPRA feedback, her correct responses improved immediately and demonstrated mastery for both curricular goals each within three sessions. During the generalization probe with a novel student, Participant E maintained high accuracy, achieving 95% for object use imitation and 100% for manipulating pegs.

In baseline under the virtual training condition, Participant E demonstrated mastery in pointing to colors within two sessions, and following vocal directions within four sessions, therefore did not require intervention. During the subsequent generalization probe, performance was maintained at 100% accuracy across both curricular goals (please see Fig. 3).

During baseline under the virtual training condition, Participant F's correct responses remained low. For the pointing-to-color program, correct responses ranged from 2 to 3 across sessions, whereas in the followable directions program, performance decreased from 5 correct responses to 1 across three baseline sessions. Following the introduction of the intervention with TPRA feedback, the participant's performance improved immediately. She reached mastery in the pointing-to-colors program within four sessions and in the followable directions program within five sessions. During the maintenance probe,

her correct responses decreased, with 80% correct responses for the pointing-to-colors program and 60% for the following vocal directions program.

During baseline under the in-person training condition, Participant F demonstrated rapid acquisition and reached mastery in both object use imitation and manipulating pegs within four sessions, therefore did not require intervention. During the subsequent generalization probe with a novel student, her correct responses remained high, with 95% correct responses for manipulating pegs and 90% for object use imitation (please see Fig. 3).

Social Validity

Our results indicated strong social validity across both participants and experimenters. Three participants responded and they rated the training procedures very favorably. All participants scored the highest possible rating (5) to the importance of systematic teacher training, the ease of learning both training procedures, the usefulness of virtual modeling and feedback, improvements in instructional accuracy, overall satisfaction with both training formats, and increased confidence in their own delivering accurate instruction. The only exception was a score of 4 by one participant for the usefulness of the in-person modeling demonstration. When asked about their preferred training format, two participants preferred the in-person format, whereas one participant indicated a preference for a hybrid approach combining virtual and in-person training.

Four experimenters responded and they rated both supervision formats favorably. All experimenters rated virtual supervision as highly effective in improving instructional accuracy (5), and three of the four experimenters gave the highest rating for overall satisfaction. The lowest ratings were for the feasibility of implementing virtual supervision within daily classroom practice, with one participant rating this item as 3 and the remaining participants rating it 4, 4, and 5. All experimenters indicated a preference for in-person supervision.

Discussion

The present study demonstrated that both virtual and in-person supervision resulted in great improvements in teachers' instructional accuracy. We found the differences between supervision modalities to be negligible. In another word, virtual supervision can be as effective as traditional in-person supervision under controlled conditions. The current study appears to be the first one that directly compared delivering supervision in person with virtually to all the components of the learn unit as well as the first to be conducted in a school setting. The study’s contribution to the literature is on teacher training by highlighting how differences in training environments may or may not produce unique contextual interventions or considerations when supervising novice behavior analysts.

An important finding of the present study is that modeled instruction alone (either in person or via video) was not sufficient to produce criterion-level responses. During baseline, participants observed demonstrations of modeled instruction and then implemented instruction independently. Despite noted

improvements, their correct responses did not reach mastery except for one participant with one instructional program. This finding suggests that modeling may not be sufficient to serve as a standalone training approach. One possible explanation is that modeling demonstrates only the correct responses and therefore cannot directly address the participant's individual teaching errors. Participant B's baseline responses in the tact picture program are a case in point. Although overall accuracy improved, the participant consistently emitted the same instructional errors across four sessions, including reinforcing incorrect responses during correction procedures, omitting an independent opportunity to respond after an error, and presenting incorrect antecedents (e.g., "what is this?"). The persistence of these errors suggests that modeling alone was unlikely to lead to mastery. Therefore, the introduction of contingent feedback was needed despite the ascending baseline trend.

Consistent with prior research, the present study demonstrated that modeling alone did not produce reliable criterion-level responding across all participants. However, some participants in the present study did achieve mastery from modeling alone. One plausible explanation could lie in dosage and distribution of modeling exposure. In prior research (e.g., Du et al., 2026), the experimenters only presented a limited number of exemplars (e.g., five learn units), whereas in the present study, participants were exposed to both video and in-person modeling at the start of each training session. This variability suggests that modeling may support acquisition under certain conditions, when prerequisites or prior learning histories are present. However, the addition of contingent feedback through the TPRA procedure resulted in consistent and rapid improvements across all participants, indicating that feedback is a critical component for establishing accurate and stable instructional performance.

In contrast, the experimenters' contingent feedback through the TPRA measure resulted in immediate and rapid increases in accurate responding across participants. These findings are consistent with previous research demonstrating the effectiveness of TPRA as a teacher training intervention in improving instructional accuracy for new teaching staff (e.g., Nuzzolo et al., 2025). The TPRA measure allowed supervisors to identify and correct errors in the participant's antecedent delivery, response recording, and consequence implementation in real time, which likely accelerated skill acquisition and promoted consistent performance at or above criterion levels. Taken together, these results suggest that regardless of the delivery approach (virtual or in person) contingent feedback appears to be necessary to achieve and maintain high levels of instructional accuracy.

An interesting question raised by the study is the possibility that difference in teaching accuracy is affected by the types of curricular goals being taught. Although all participants received the same training procedures, their performance varied across goals. Some goals appeared easier to implement accurately, while others required additional feedback and practice to reach mastery. These differences may be related to the participants' prior instructional histories, familiarity with certain program types, or the complexity of the selected curricular domains. For example, goals requiring more precise correction procedures may place greater demands on teacher performance than simpler instructional formats.

Some serendipitous findings, of three of the six participants, S, K and B, should be noted in their second set of training repertoires whether they were in-person or virtually supervised. These three participants did not require the TPRA measure, or any supervision to learn the novel teaching repertoires as compared to participants T, N and M. they already demonstrated transformations whereas the others had to be taught. The data show learning occurred without any intervention and the results also generalized to a novel student in all cases. A possible explanation for this learning without any explicit training or instruction could be the transformation of stimulus function (TSF) (Dymond & Rehfeldt, 2000) of novel learn unit repertoires from one supervision style to the next. TSF involves the acquisition of untaught responses to a stimulus which previously evoked only a single taught response (Eby et al. 2010). One possible behavioral interpretation is TSF. Previous studies have demonstrated the emergence of untaught responses through TSF across various responses, including written and vocal spelling (Eby et al. 2010; Greer et al., 2005), autoclitic frames (Luke et al., 2011; Speckman et al. 2012), mands and tacts (Nuzzolo & Greer, 2004) and math repertoires (Sun & Greer, 2022) in the behavioral literature for typical and atypical children and it is possibly the explanation for the emergence of the untaught repertoires in this experiment as well.

Social validity findings suggest that both training formats were viewed as practical and effective and that the procedures improved new teaching staff's confidence. It is also important to note that additional challenges may arise in virtual learning contexts due to ongoing technological advancements. For example, recently added Zoom features (e.g., automatic speaker-focused zoom) may limit the supervisor's observation of the child or the teacher when they are not talking.

Our data also show that both participants and supervisors appeared to show more preference toward in-person training. This face to face approach allows supervisors to model correct responses in real time and provide immediate support for behavior management. This may be especially important for less experienced staff who are still learning to developing instructional control. BCBA's and BCBA-D's who supervise instructors via telehealth report that a major concern when delivering supervision is the ability to model and demonstrate strategies to their supervisee (Sipila-Thomas & Brodhead, 2024). When the model and support are virtual, it seems to be much more difficult for both the supervisor and supervisee to work together when a student is emitting challenging behavior. In contrast, live supervision enables supervisors to intervene promptly when challenging behaviors occur. However, in the current study, we controlled these factors to ensure comparability between in-person and virtual conditions.

The comparison between virtual and in-person training highlights several practical advantages and limitations of each approach. Virtual training offers clear logistical benefits, including reduced labor costs, elimination of commute time, and increased accessibility for both supervisors and trainees. With the need for behavior analytic services rapidly growing (Thompson et al., 2026), virtual supervision offers an effective way for one supervisor to provide more supervision reducing factors such as commuting, limited hours and higher costs. However, social validity surveys showed that it was generally less preferred by trainees. In addition, it required careful setup to be effective. Challenges such as inconsistent audio quality, limited video clarity, and camera angles sometimes restricted the supervisor's

ability to observe the entire interactions between the trainee and the student during the instructional sessions.

We controlled environmental variables by conducting both in-person and virtual sessions in the same research room. This addressed a limitation of previous research (Peysin et al., 2022), in which virtual sessions typically occurred in the home whereas in-person sessions occurred in school settings, potentially confounding comparisons. Although this research was conducted in a school, it was not done in a classroom setting which would have included other students and teachers separate than the target participants. This, in and of itself, would also seem to be a limitation. Many behavior analysts serve as consultants in public and private school settings and some in a virtual capacity. Collecting data in this setting would have further served to help test whether virtual supervision, including many distractors, especially auditory, is feasible or at least as effective as in person. All of the collective research done on this topic has been done in a very controlled setting, and this is something the authors are considering for future research. Furthermore, all the students chosen for the participants to work with did not have a history of interfering behaviors during instruction. We do realize this may not be representative of cases BCBA's sometimes work with in school settings, but it was a variable we felt needed to be controlled for to acquire results for the dependent variable.

One limitation is the different baseline sessions across participants and conditions. Baseline phases ranged from three to five sessions. In most cases, the number of baseline sessions was held constant across the two training conditions for each participant. However, for Participant D, baseline for tracing marks was terminated after three sessions due to consistently zero correct responding. At the same time, baseline continued for the second program, during which the participant demonstrated improvement and ultimately reached mastery without intervention. This resulted in an unequal number of baseline sessions within that participant's conditions. More broadly, baseline termination followed a decision rule based on data patterns: phases were ended after three sessions when the participants' responding was low and stable (or decreasing), and we extended to five sessions when stability had not yet been demonstrated. Although this approach was consistent with single case design requirement, the variability in baseline length introduces a limitation and should be considered when interpreting the results.

Declarations

Author Contribution

A - Concept for the work, writing and reviewing/editing manuscriptB - data collection, graphing, design and writing and reviewingC - data collection, IOA, graphingD - data collection, IOA, graphing, editingE - data collection, IOAF - concept, design and edits

Data Availability

The data that support the findings of this study are stored at the setting the research took place and available from the corresponding author upon reasonable request.

References

1. Akers, J.S., Davis, T.N., McGinnis, K. & Swensson, R.M. (2024). Effectiveness of remote delayed performance feedback on accurate implementation of caregiver coaching.*Journal of Behavioral Education,* 33, 320-343. doi.org/10.1007/s10864-022-09487-0
2. Andzik, N.R. & Kranak, M.P. (2021). The softer side of supervision: Recommendations when teaching and evaluating behavior-analytic professionalism. *Behavior Analysis in Practice*, 21, 65-74.
3. American Telemedicine Association. (2017). Telehealth: Defining 21st century care.https://www.americantelemed.org/resource/why-telemedicine
4. Behavior Analyst Certification Board (2018). *A summary of ethics violations and code enforcement activities.* https://www.bacb.com/wp-content/ethics-violations-and-code-enforcement-activities
5. Britton, L. N., & Cicoria, M. J. (2019). *Remote fieldwork supervision for BCBA® trainees.* Elsevier Academic Press.
6. Chen, S., Wathen, C. & Speciale, M. (2020). Online clinical training in the virtual remote Environment: Challenges, opportunities and solutions. *The Professional Counselor,* 10(1), 78-91.
7. Dymond, S. & Rehfeldt, R.A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst,* 23, 239-254. doi: 10.1007/BF03392013
8. Eby, C. Greer, R.D., Tullo, L.D., Baker, K.A. & Pauly, R. (2010). Effects of multiple exemplar instruction on transformation of stimulus function across written and vocal spelling responses by students with autism. *The Journal of Speech and Language Pathology*, 5(1), 20-31 doi.org/10.1037/h0100262
9. Figge, C.J., Kane, J.C., Skavenski, S., Haroz, E., Mwenge, M., Mulemba, S., Aldridge, L.R., Vinikoor, M.J., Sharma, A., Inoue, S., Paul, R., Simeda, F., Metz, K, Bolton, C., Kemp, C., Bosomprah, S., Sikazwe, I. & Murray, L.K. (2022). Comparative effectiveness of in- person vs. remote delivery of the common elements treatment approach for addressing mental and behavioral health problems among adolescents and young adults in Zambia: protocol of a three-arm randomized controlled trial. *Trials*, https://doi.org/10.1186/s13063-022-06319-4
10. Greer, R.D. Yaun, L. & Gautreaux, G. (2005). Novel dictation and intraverbal responses as a function of multiple exemplar instructional history. *The Analysis of Verbal Behavior,* 21(1), 99-116. DOI: 10.1007/BF03393012
11. Greer, R. D., & Hogin-McDonough, S. (1999). Is the learn unit the fundamental unit of pedagogy? *The Behavior Analyst, 20,* 5-16.
12. Hartzler, B., Hinde, J., Lang, S., Correia, N., Yermash, J., Yap, K., Murphy, C.M., Ruwala, R., Rash, C.J., Becker, S.J. & Garner, B.R. (2022). Virtual training is more cost effective than in-person training for preparing staff to implement contingency management. *Journal of Technology in Behavioral Science*, 8, 225-264.

13. Ivanova, J., Shah, B., Foote, C. & Cummins, M.R. (2025). Assessing the costs and savings of telemedicine: Insights from the consolidated framework for implementation research.*Digital Health*, doi:10.1177/20552076251314552
14. Lerman, D.C., O'Brien, M.J., Neely, L., Call, N.A., Tsami, L., Schieltz, K., Berg, W., Graber, J., Huang, P., Kopelman, T. & Cooper-Brown, L.J. (2020). Remote coaching of caregivers via telehealth: Challenges and potential solutions. *Journal of Behavioral Education*, 29(2), 195-221.
15. Lindgren, S, Wacker, D., Suess, A., Schieltz, K., Pelzel, K., Kopelman, T., Lee, J., Romani, P, & Waldron, D. (2016). Telehealth and autism: Treating challenging behavior at lower cost. *Pediatrics*, 137, S167-S175.
16. Luke, N., Greer, R.D., Singer-Dudek, J. & Keohane, D. D. (2011). The emergence of autoclitic frames in atypically and typically developing children as a function of multiple exemplar instruction. *The Analysis of Verbal Behavior,* 27, 141-156.
17. Nuzzolo, R., Du. L., Buttigieg, S., Greer, R. D. & Pino, J. (2025). *The Teacher Performance Rate and Accuracy Measure as a Teacher Training Intervention. Journal of Behavioral Education.* https://doi.org/10.1007/s10864-025-09592-w
18. Nuzzolo, R. & Greer, R.D. (2004). Emergence of untaught mands or tacts of novel adjective-object pairs as a function of instructional history. *The Analysis of Verbal Behavior,* doi:10.1007/BF03392995
19. Pantermuehl, R.M. & Lechago, S.A. (2015). A comparison of feedback provided in vivo versus an online platform on the treatment integrity of staff working with children with autism. *Behavior Analysis in Practice,* 8, 219-222. doi: 10.1007/s40617-015-0059-y
20. Perle, J.C. & Zheng, W. (2023). A primer for understanding and utilizing telesupervision with With healthcare trainees. *Journal of Technology in Behavioral Science*, doi: 10.1007/s41347-023-00322-5
21. Peysin, J., Fienup, D., & Romas, S. (2022). An Experimental Comparison of In-Person and Remote Instruction for Preschoolers with Disabilities. *Behavior analysis in practice*, *16*(2), 558–572. https://doi.org/10.1007/s40617-022-00737-1
22. Robertson, J., Thomson, K., Hume, M., Magnacca, C., Marcinkiewicz, A. & Cauchi, J (2023). Evaluation of telehealth training for behavior technicians to implement common behavioral protocols. *Journal of Behavioral Education,* 34, 53-79. doi.org/10.1007/s10864-023-09521-9
23. Ross, D. E., Singer-Dudek, J., & Greer, R. D. (2005). The teacher performance rate and accuracy scale (TPRA): Training as evaluation. *Education and Training in Developmental Disabilities, 40(4),* 411-423. http://www.jstor.org/stable/23879958
24. Simmons, C.A., Ford, K.R., Salvatore, G.L. & Moretti, A.E. (2021). Acceptability and feasibility of virtual behavior analysis supervision. *Behavior Analysis in Practice*, 14(4), 927-943.
25. Singer-Dudek, J., Du, L., Greer, A., Dakopolos, A., Jahromi, L.B., Brassard, M.R., & Greer, R.D. (2023). The Verbal Behavior Development Assessment-Revised: Convergent and Divergent Validity in a Sample of Preschoolers with Autism Spectrum Disorder. *Journal of Developmental and Physical Disabilities*, 1-19.

26. Sipila-Thomas, E.S. & Brodhead, M.T. (2023). A survey of barriers experienced while providing Supervision via telehealth:implications for future research and practice. *Behavior Analysis and Practice*, 17(1), 70-86.
27. Speckman, J., Greer, R.D. & Rivera-Valdes, C. (2012). Multiple exemplar instruction and the emergence of generative production of suffixes as autoclitic frames. *The Analysis of Verbal Behavior,* 28(1), 83-99. https://doi.org/10.1007/
28. Sullivan, A.B., Davin, S.A., Lapin, B., Schuster, A.T., Dweik, R.A., Murray, K.F., Rehm, S.J. & Machado, A.G. (2022). Effects of flexible scheduling and virtual visits on burnout for clinicians. *Multiple Sclerosis and Related Disorders*, 60 https://doi.org/10.1016/j.msard.2022.103705
29. Sun, Y. & Greer, R.D. (2022) Stimulus control for making math verbal. *The Psychological Record,* 73(1), 33-52. doi 10.1007/s40732-022-00531-z
30. Thompson, R.L., McMahon, J.S., Herrara, A. A. & Colon, C.L. (2026). Developing a comprehensive supervision program for aspiring behavior analysts: A tutorial. *Behavior Analysis in Practice,* 19, 829-844. doi.org/10.1007/s40617-025-01153-x
31. Wilczenski, F.L. & Coomey, S.M. (2005). Cyber-communication: Finding it's place in school counseling practice, education and professional development. *Professional School Counseling*, 9(4) doi 10.1177/2156759X0500900404
32. Wood, J.A.V., Miller, T.W. & Hargrove, D.S. (2005). Clinical supervision in rural settings: A telehealth model. *Professional Psychology*: *Research and Practice*, 36(2), 173-179.

Tables

Table 1.

Participants' demographics

Dyad	Participant	Gender	Age	Education	Ethnicity
1	A	F	22	BS	Black/Hispanic
	B	F	22	HS	Hispanic
2	C	F	23	HS	Hispanic
	D	F	22	HS	Hispanic
3	E	F	32	BA	Hispanic
	F	F	22	HS	Hispanic

Table 2.

Sequence of Training Condition and Instructional Program

		First Round		Second Round	
		Virtual	In person	Virtual	In person
Dyad 1	Participant A	Tact pictures Follow directions			Trace symbols Gross motor imitation
	Participant B		Tact pictures Follow directions	Trace symbols Gross motor imitation	
Dyad 2	Participant C		Trace symbols Gross motor imitation	Tact pictures Follow directions	
	Participant D	Trace symbols Gross motor imitation			Tact pictures Follow directions
Dyad 3	Participant E		Object use imitation Manipulate pegs	Follow directions Point to colors	
	Participant F	Object use imitation Manipulate pegs			Follow directions Point to colors

Table 3.

Learn units to criterion by condition

	Virtual			In person		
A	2	3	5	2	3	5
B	5	5	10	4	6	10
C	3	4	7	4	5	9
D	6	NA	6	4	2	6
E	2	4	6	3	3	6
F	4	5	9	NA	NA	
Total			43			36

Table 4.

Interobserver agreement collected during both conditions

Virtual			In person	
% sessions with IOA, average IOA% (range)				
Participant	Baseline	Intervention	Baseline	Intervention
A	70%, 100%	100%, 100%	70%, 100%	100%, 96.25% (90-100%)
B	40%, 100%	83%, 99% (95-100%)	40%, 100%	40%, 100%
C	66%, 97.5% (95-100%)	75%, 100%	33%, 100%	33%, 100%
D	33%, 100%	50%, 100%	33%, 100%	
E	50%, 100%		40%, 100%	50%, 100%
F	33%, 100%	16%,100%	33%, 100%	
Total	53%, 99.7%, (95-100%)	66%, 99.6% (95-100%)	41%, 100%	42%, 98% (90-100%)

Note. *A black cell indicates the participant was not exposed to this treatment condition due to their performance in baseline.*

Table 5.

Treatment fidelity agreement collected during both conditions

Virtual			In person	
% sessions with treatment fidelity, fidelity %				
Participant	Baseline	Intervention	Baseline	Intervention
A	100%, 100%	100%, 100%	70%, 100%	100%, 100%
B	40%, 100%	83%, 100%	40%, 100%	40%, 100%
C	66%, 100%	75%, 100%	33%, 100%	33%, 100%
D	50%, 100%	50%, 100%	33%, 100%	
E	50%, 100%		40%, 100%	50%, 100%
F	66%, 100%	16%,100%	33%, 100%	
Total	61%, 100%	33%, 100%	41%, 100%	42%, 100%

Note. *A black cell indicates the participant was not exposed to this treatment condition due to their performance in baseline.*

Figures

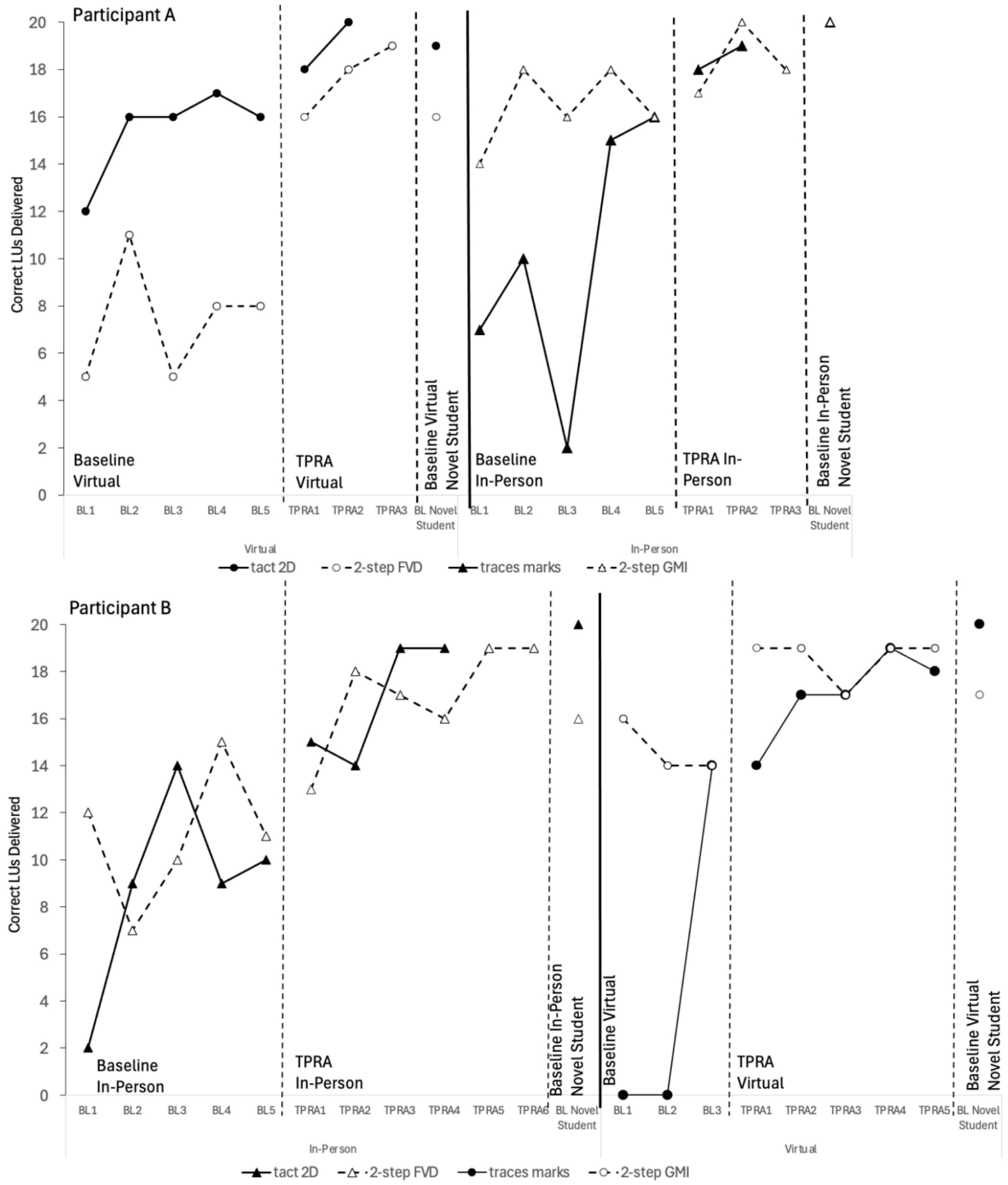


Figure 1

Legend not included with this version

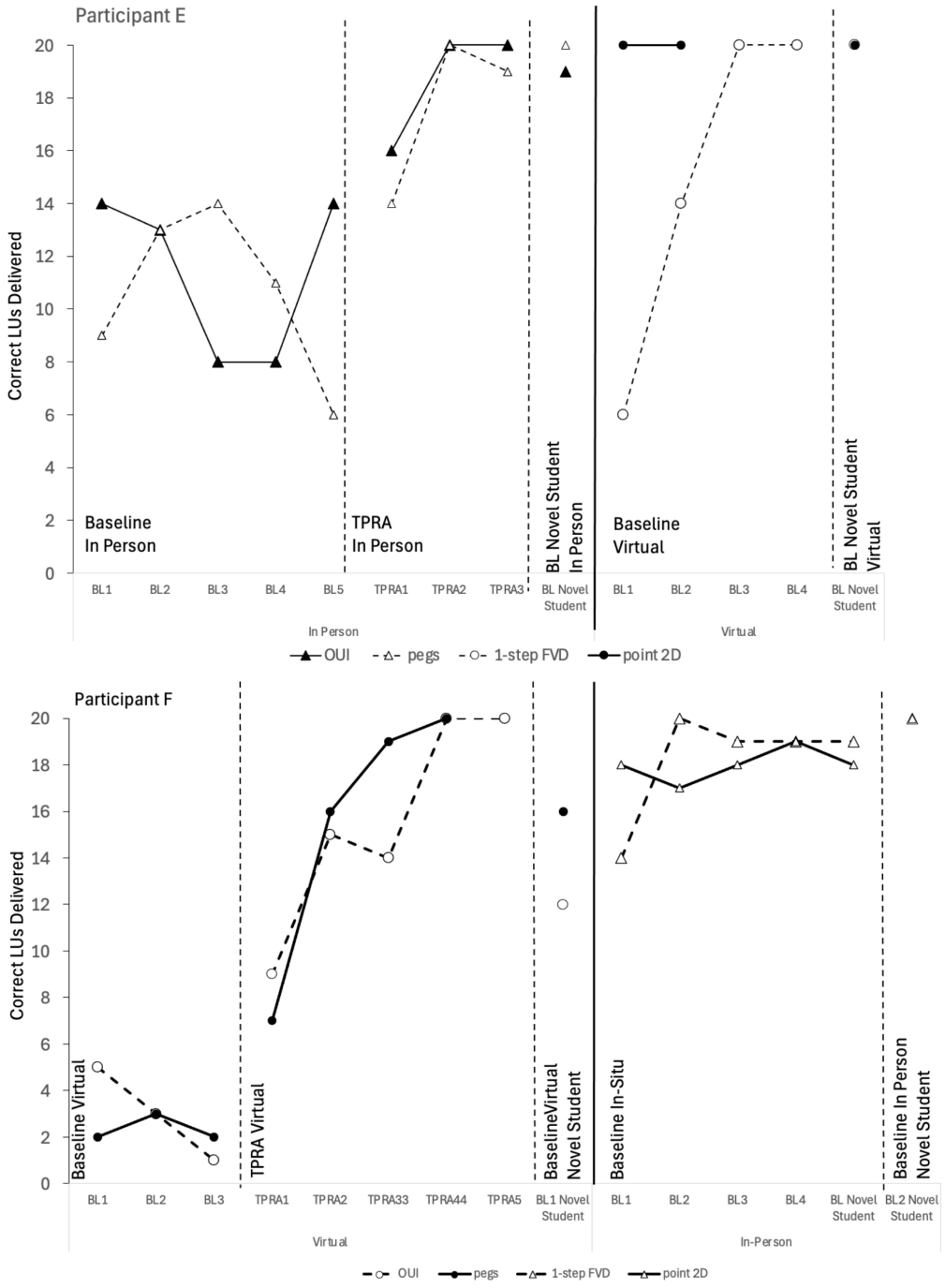


Figure 2

Legend not included with this version

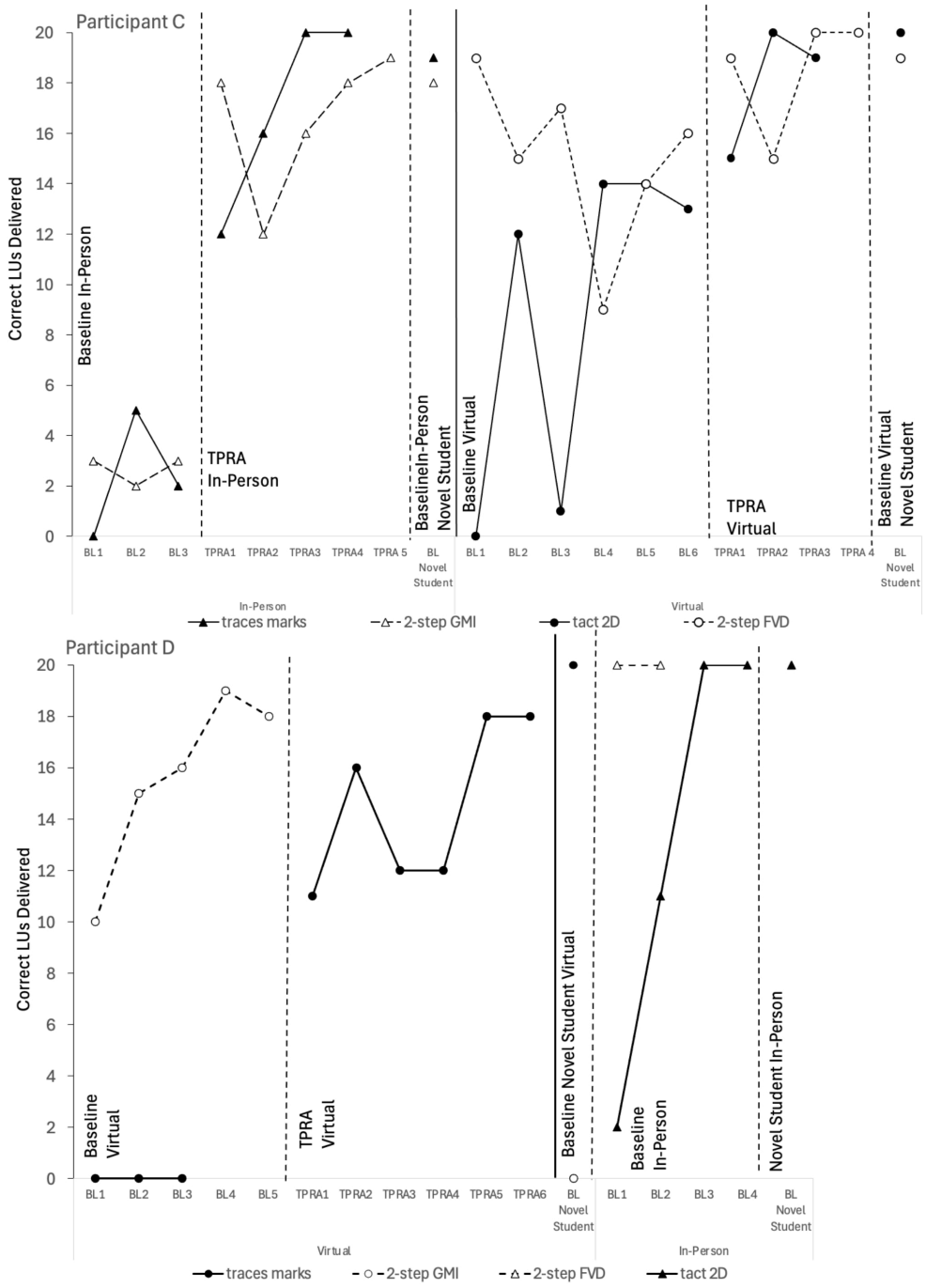


Figure 3

Legend not included with this version

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixV.docx