An In-Depth Analysis of the Performance Diagnostic Checklist – Human Services (PDC-HS)

By

Meagan Harder
B.A., Elementary Education and Special Education, 2015

Submitted to the graduate degree program in Applied Behavioral Science Department and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

Chair: Robin Kuhn

Florence DiGennaro Reed

Jean Hall

Date Defended: 7/15/2020

The thesis committee for Meagan Harder certifies that this is the approved version of the following thesis:

An In-Depth Analysis of the Performance Diagnostic Checklist – Human Services (PDC-HS)

Chair: Robin Kuhn

Date Approved: 7/15/2020

Abstract

The current study aims to assess a digital version of the Performance Diagnostic Checklist – Human Services (PDC-HS; Carr et al., 2013; Carr & Wilder, 2016). In Experiment 1, within three non-concurrent multiple baseline designs, three groups of participants (i.e., a Traditional PDC-HS group, a Baseline Control group, and a Digital PDC-HS group) were tasked with solving a 3x3 sliding puzzle across three conditions. Participants in the Traditional and Digital PDC-HS groups experienced baseline, domain manipulation, and barrier removal conditions. Participants in the Baseline Control group experienced baseline contingencies across all three conditions. During the baseline condition, participants were presented with an online version of a 3x3 sliding puzzle and simply asked to solve it. During the domain manipulation condition, participants were asked to solve the puzzle again, however, some barriers were removed to promote fluent puzzle solving. During the barrier removal condition, participants were asked to solve the puzzle again immediately after receiving behavioral skills training on how to solve the puzzle fluently. All participants were administered the PDC-HS after completing the first and second conditions; participants in the Traditional PDC-HS group were administered the PDC-HS via interview and participants in the Digital PDC-HS and Baseline Control groups were administered the PDC-HS via Qualtrics® Research. Participants in Experiment 2 experienced the same procedures as Experiment 1; however, modifications were made to the barriers removed in the domain manipulations condition. In both Experiments 1 and 2, the majority of participants in both the Traditional and Digital PDC-HS groups only solved the puzzle following behavioral skills training, suggesting lack of training was the barrier to fluent task completion. The results of Experiment 1 indicated that while both the Digital and Traditional PDC-HS identified *Training* as barrier, other domains were identified as well. The results of Experiment 2 suggest that the Traditional PDC-HS may accurately identify barriers within a manufactured work environment. Taken together, these findings, and the viability of a Digital PDC-HS, are discussed in the context of existing literature as well as next steps toward improved validation of the PDC-HS.

Acknowledgments

First, thank you so much to my family and friends for their love, support, and words of encouragement. I appreciate your understanding and love as my project ended. While my amount of available time decreased, your love and support never did. Specifically, thank you to my amazing husband who continued to help make our home a happy and loving place. Thank you for waiting to eat dinner with me until 8 PM, meeting me in the student union for dinner, and waiting to walk our dog until 10 PM, just so that we could share a few precious moments together.

Thank you to Dr. Robin Kuhn for teaching me invaluable lessons inside and outside of the classroom. Thank you so much to my committee members, Drs. Jean Hall, Florence D. DiGennaro Reed, and Robin Kuhn for dedicating your time and expertise in considering this project. This study received immeasurable support from Alyssa Haase's regular counsel and responsiveness and gained traction during the KU's thesis accelerator program. Thank you to Rebecca Woolbert, Morgan Benson, Daryl Stewart, Jessika Tucker, Kylie Baima, and Sammie Kruse for being willing to score and help interpret the data. Your dedication outside of your own schooling and work duties will always be remembered and appreciated.

Finally, thank you to my student who loves bubbles, butterflies, and yellow who set me on this path and thank you to her amazing family who introduced me to the wonderful world of Applied Behavior Analysis.

Table of Contents

List of Figures

List of Tables

Introduction and Literature Review

Applied behavior analysis has been implemented within organizational settings in different ways, including analyzing systems (Bucklin et al., 2000), promoting safety (McSween, 2003), and managing employees (Austin, 2000). Performance Management is a sub-area within organizational behavior management (OBM) specializing in the analysis of antecedents and consequences influencing employee behavior, and the development of interventions to change maintaining variables for the purpose of improving employee performance (Austin, 2000). To help identify the maintaining contingencies and guide interventions, an assessment may be used.

Performance Assessments

Typically, assessments in the discipline of behavior analysis involve the identification of environmental variables maintaining behavior. However, assessments in OBM tend to focus on an individual, process, or systems level analysis (Austin, 2000). Target behaviors, as well as intervention and assessment procedures, can vary by level. A systems level analysis examines company-wide consequences such as profits, competitors, consumer services, expenses, and ultimately helping companies achieve their mission (Abernathy, 2014). The process level analysis reviews how processes or work are performed within an organization (Austin, 2000), including design, development, sales, and delivery of products and services (Sasson et al., 2006). Performance level reviews specifically assess how the employee is performing and how well they are completing their work (Austin, 2000).

Generally, assessments in behavior analysis involve examining variables maintaining behavior at the three levels described above. Assessments designed to measure staff behavior can be organized by three different methods: 1) experimental, 2) descriptive, and 3) informant (Austin, 2000). Experimental analyses involve manipulating consequences to identify the variables contributing to performance problems (Iwata et al., 1982/1994). While experimental analyses can be common in clinical settings, an example being functional analysis (FA; Iwata et al., 1982/94), very few studies demonstrate their use in the OBM literature (Wilder & Hodges, 2018). However, while both the *antecedents* and *consequences* of a behavior are manipulated in FAs, past experimental methods in organizational settings have manipulated only *antecedents*, perhaps due to the difficulty of manipulating consequences in a workplace setting (Wilder & Hodges, 2018) an example being Pampino et al. (2005). The investigators examined the antecedents that possibly contributed to the lack of daily submission of billing information. Three possible antecedents were hypothesized to contribute to inadequate employee performance: a) inaccurate remembering of 4-digit operation codes; b) a deficit in locating and transferring 4-digit codes into a spreadsheet; and c) inaccurate typing of code numbers into the spreadsheet on a laptop. An experimental analysis including three different assessments was conducted to determine the antecedent contributing to the inadequate performance. The first assessment was designed to examine the inability to remember the four digit codes. Participants were given a list of operations that corresponded with codes they had to recall from memory. The second assessment involved shifting the codes to an alternate spreadsheet to determine if errors were due to transferring information. The final assessment examined a barrier to fluency by asking participants to input codes on a laptop to determine if it was the computer as opposed to ability. The results suggested that product knowledge (Assessment A) and data entry (Assessment C) contributed to inaccurate data entry. Training was completed to reteach product knowledge as well as data entry on the laptop. After training had been completed, accurate submission of billing information increased, and inaccurate submission decreased.

In contrast to experimental methods, descriptive assessments obtain information simply through direct observation as opposed to variable manipulations. When conducting a descriptive assessment, employee behavior is observed, and antecedent and consequence events are recorded and analyzed. Fante et al. (2010) completed a descriptive assessment to determine the antecedents of inaccurate wrist posture among pharmacy technicians. A descriptive assessment was completed to determine if an environmental variable introduced by participants, a box to support wrists, actually contributed to wrist safety. Observers collected data on wrist posture and support when the box was present and when it was not. The results suggested that when participants were using the box, wrist safety increased substantially. Based on the results, experimenters consulted with an occupational and physical therapist and obtained a keyboard that provided wrist support. The results suggested that wrist support and safety improved when the keyboard was used.

Finally, informant or indirect assessments are the most common method applied within organizations (Wilder & Hodges, 2018). Informant assessments are typically conducted via interview or self-report. Employees are typically asked questions to help determine if performance barriers are skill deficits or other environmental variables, such as lack of consequences, inefficient processes, or absent materials. Different versions of informant assessments exist, such as general assessments like guiding questions (Mager & Pipe, 1984) and specific assessments such as the PIC/NIC (Daniels & Bailey, 2014) or the Performance Diagnostic Checklist (PDC; Austin, 2000).

In the workplace and everywhere else, “the organism is always right” (Skinner, 1977, p. 1007), meaning, in simplified terms, that organisms behave in accordance with contingencies operating in the environment. However, “misarranged stimulus control” can be interpreted as errors, which may occur within at any part of the three term contingency – the antecedent, the behavior, or the consequence (Fisher et al., 2014). The PDC, created by Austin in 2000, helps identify the antecedents and consequences governing employee performance. To create the PDC, managers and consultants were interviewed to determine common questions asked for problem solving behaviors. Questions were divided into four categories: Antecedents and Information; Equipment and Processes; Knowledge and Skills; and Consequence. The PDC has been used in the food service industry (Austin et al., 2005; Amigo et al., 2008; Pampino et al., 2004); stores (Doll et al., 2007; Eikenhout & Austin, 2005; Loughrey et al., 2013); schools (Hybza et al., 2013) and clinics (Gravina et al., 2008, Lebbon et al., 2011). The PDC has been used to assess cleaning tasks (Amigo et al., 2008; Austin et al., 2005; Doll et al., 2007; Pamipino et al., 2008), complex behaviors (Lebbon et al., 2011; Hybza et al., 2013) and social behaviors (Eikenhout & Austin, 2005; Loughrey et al., 2013).

Despite its versatility, a modified version of the PDC was created specifically for human service settings (Carr et al., 2013; Carr & Wilder, 2016). The PDC was primarily designed for the organizational setting, and as such, it includes items that were not relevant to the human-service setting, such as *Equipment and Processes*. In the *Equipment and Processes* section of the PDC, the following questions are posed: “If equipment is required, is it reliable? Is it in good working order? Is it ergonomically correct?” (Austin, 2000, p. 340). While the human service settings may occasionally have equipment, “the presence of irrelevant items or items that need to be translated into more contextually relevant terms might diminish the instrument’s utility” (Carr et al., 2013, p. 18).

PDC-HS

To create the Performance Diagnostic Checklist – Human Services (PDC-HS), the PDC was applied to common problems in the human service setting, such as inaccurate data collection, poor treatment implementation, inadequate development of program materials, poor attendance and tardiness, failure to report problems to supervisors, and poor graph construction (Carr et al., 2013). The problem-solving process identified the following areas for revision: section titles, section order, question wording, and question order. After receiving input on the wording from 11 behavior analysts, the PDC-HS was piloted and then applied to a target behavior.

The PDC-HS has the same format as the PDC; twenty questions are answered indirectly by supervisors or staff as well as direct observation conducted by researchers or supervisors, with questions covering four domains: (a) Training, (b) Task Clarification and Prompting, (c) Resources, Materials, and Processes, and (d) Performance Consequences, Effort, and Competition (Carr et al., 2013; Carr & Wilder, 2016). Similar to other informant styles, the PDC-HS is conducted by interview and direct observation (Carr et al., 2013; Wilder & Hodges, 2018)There are between four and six questions in each domain; of all twenty questions, thirteen may be answered using informant report and seven of which should be answered from direct observation. Each question that is scored "No" on the PDC-HS is interpreted as an opportunity for intervention. Priority is given to domains in which multiple items are endorsed (Carr & Wilder, 2016).

The purpose of the original study conducted with the PDC-HS was to assess its effectiveness at identifying performance barriers of staff working within an early intensive behavioral intervention center, as well as to examine its predictive validity. Predictive validity specifies the extent to which assessments have accurate outcomes, or this case, accurately identify barriers (Cronbach & Meehl, 1955). To assess the predictive validity of the PDC-HS, a nonfunction-based and then a function-based intervention was applied. The non-function based intervention was not expected to produce a meaningful behavior change. The indicated intervention was then applied to assess if it resulted in meaningful behavior change. Carr et al. (2013) targeted room cleanliness across eight rooms, which was measured by the percentage of tasks participants completed on a cleanliness checklist. Three behavior analysts were interviewed about employee performance. The behavior analysts responding on the PDC-HS identified lack of training and feedback as potential barriers. Participants in six out of eight rooms received baseline and the indicated interventions. Participants in two out of the eight rooms received baseline, non-indicated interventions, and an indicated intervention. The results suggested that for the rooms with participants who only experienced baseline and the indicated interventions, the percent of tasks completed on the cleanliness checklist not only immediately improved relative to baseline, but also remained stable at a high level for the remainder of the study. In the two rooms with participants who received the non-indicated interventions, the participants in one room increased its cleanliness by 11% relative to baseline, however, it was not an effective intervention as the cleanliness criterion was not met. In the other room, participants who received a nonindicated intervention experienced a 6% decrease of the amount of the checklist completed, thus during the nonindicated intervention participants cleaned rooms less thoroughly relative to baseline. The combination of the indicated intervention producing positive behavior change and the non-indicated intervention producing negative or ineffective behavior change validates the PDC-HS.

In 2016, Carr and Wilder revised the PDC-HS and rephrased three questions. Item five in *Resources, Materials, and Processes* and items four and five in *Performance Consequences,* *Effort, and Competition* were worded such that a response of "Yes" indicated an opportunity for intervention in the original publication. The three questions were reworded so that an answer of "No" (instead of "Yes") indicated an opportunity for intervention.

Since its development, the PDC-HS has been used to assess a variety of staff performance issues such as room cleanliness (Carr et al., 2013); closing doors (Ditzian et al., 2015); rate of mand, tact, and intraverbal opportunities (Wilder et al., 2018), and error correction procedures (Bowe & Sellers, 2018); vocational responsibilities (Smith & Wilder, 2018; Hess, 2019); and tardiness (Merritt et al., 2019). Preliminary studies have also been completed to assess the validity and reliability of the instrument (Wilder et al., 2019). The PDC-HS has also been used across various human service settings, including educational environments (Bowe & Sellers, 2018; Merritt et al., 2019), clinical placements (Carr et al., 2013; Ditzian et al., 2015; Wilder et al., 2018), thrift stores (Smith & Wilder; 2018), and in a library (Hess, 2019). The PDC-HS has assessed performance challenges faced by individuals serving different roles, for example paraprofessionals (Carr et al., 2013; Merritt et al., 2019), therapists (Wilder et al., 2018), employees in non-educational settings (Ditzian et al., 2015); and employees diagnosed with intellectual disabilities (Hess, 2019; Smith & Wilder, 2018). Two additional versions of the PDC have also been created, the PDC-Safety (Martinez-Onstott et al., 2016) and the PDC-Parent (Villacorta, 2017), indicating the perceived value of the PDC assessment across disciplines.

All domains of the PDC-HS have been assessed. Two studies found *Training* to be the only primary barrier (Bowe & Sellers, 2018; Smith & Wilder, 2018) and one found *Performance Consequences, Effort, and Competition* (Ditzian et al., 2015) to be the primary barrier. All other studies resulted in a packaged intervention such as *Training* and *Performance Consequences, Effort, and Competition* (Carr et al., 2013); *Task Clarification and Prompting, Resources,* *Materials, and Processes*, as well as *Performance Consequences, Effort, and Competition* (Wilder et al., 2018; Merritt et al., 2019); and personalized packages for participants (Wilder et al., 2019).

Recently, Wilder et al. (2019) assessed the reliability and the validity of the PDC-HS by having participants watch and score videos of a performance problem twice, two to four weeks apart. Three video vignettes were created that showed three different work problems, with a different barrier presented in each. Participants not only accurately identified the barriers in their responses to questions on the PDC-HS, but they also duplicated their answers two weeks later after watching the video again. Additionally, the researchers compared participant responses within each domain and found that there was interobserver agreement (IOA) between the participants. IOA was at least 93% across all three videos and domains, with overall IOA above 95%. While the video only mimicked a work environment, it demonstrated a preliminary first step toward assessing the reliability and validity of the PDC-HS.

In a recently published brief review, Wilder et al. (2020) suggested some areas of future research on the PDC-HS. Wilder and colleagues suggested studying reliability and validity of PDC-HS responses between multiple observers in a naturalistic setting. Another opportunity for future research they described involves having employees examine their own task completion using the PDC-HS, as current procedures typically involve having a third party interview supervisors. However, given the past success of two supervisors, who were not trained in behavior analysis, interpreting the results of a PDC-HS administered by a third party and implementing an indicated intervention (Bowe & Sellers, 2018), it is possible that supervisors alone may complete the PDC-HS, and successfully select and implement interventions entirely independently despite their previous exposure, or lack thereof, to behavior analysis. By evaluating self-reported employee responses on the PDC-HS, supervisors may gain insight to barriers to performance that otherwise may not be clear (i.e., lack of childcare, traffic, etc; F. DiGennaro Reed, personal communication, July 15, 2020).

Self-Report and Signal Detection

While self-report may be useful to help identify barriers to performance, it is critical to corroborate self-report when it is used as a primary source of data. Data recorded during direct observations are measured objectively, however, self-report relies on the participant's interpretation of the events (Critchfield et al., 1998). Furthermore, self-report of private events is more difficult to authenticate because most private events are difficult to verify (Critchfield et al., 1998).

Signal Detection (SD) was originally used to quantify the extent to which participants attended to a stimulus present in their environment (e.g., light or tone; Beal & Eubanks, 2003), however, it has also been used to study self-report (Critchfield, 1993; Lerman et al., 2010; Smith et al., 2013). There are four possible mutually exclusive outcomes in SD: 1) true positive or hit, 2) true negative or correct rejection, 3) false positive or false alarm, or 4) false negative or miss (Treat & Viken, 2012). These four outcomes can be used to calculate sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV), which can then be used to estimate the utility of a standardized assessment or diagnostic measure.

Sensitivity and specificity are valuable measures to individuals who are *creating* the diagnostic measure, where as PPV and NPV are valuable measures for individuals who are *using* the measure (Treat & Viken, 2012). Sensitivity and specificity are used to quantify the precision of the *assessment*. For example, if a disease is present, the measures quantify how likely it is that the diagnostic assessment will be able to find it. PPV and NPV measure the accuracy of a positive or negative *result*. Therefore, as a *user* of the diagnostic assessment, it would be valuable to know how likely it is that the assessment yielded an accurate measure.

There are various applications of SD in the behavioral literature. Holland (1958) examined detection of a signal as a possible reinforcer and how reinforcement schedules affected responding. Bonnell et al. (2003) studied enhanced pitch sensitivity in individuals diagnosed with autism using SD. Poling et al. (2011) conducted an SD analysis to examine the technical adequacy of African Pouched Rats detection of tuberculosis.

More relevant to the present study, Critchfield (1993) examined the reliability of self-report using signal detection (SD). Critchfield asked participants to complete a delayed match-to-sample task. After participants completed the task, they were asked if they thought they scored a point by completing the task accurately. The results suggested that as more feedback was provided, participant positive self-report bias increased, indicating that participants perceived their accuracy to increase. SD may be a useful method to determine the contingencies maintaining verbal self-report (i.e., stimulus control). Furthermore, bias and discriminability may be influenced by different characteristics, ones that may not have been detected by a simple accuracy measurement.

Lerman and colleagues (2010) examined observer accuracy using an SD framework, assessing variables that may affect the specificity and bias of observer accuracy. Three experiments were completed in the study where participants watched video vignettes involving a teacher placing a demand on a student, both of whom were actors displaying varying ambiguity of a target behavior.

The purpose of Experiment 1 was to determine the effects of feedback on scoring accuracy, and more specifically, response bias. Response bias was defined as “the proclivity of the observer to judge in one direction as opposed to the other (e.g., to indicate that the signal is present rather than absent)" (Lerman et al., 2010, p. 196). Investigators were able to bias responding by either providing points for every identification of behavior or taking away points for every inaccurate recording of behavior. Experiment 2 focused on the specificity of the operational definition. The results suggested that a more specific definition produced increased accuracy of data collection relative to Experiment 1. Finally, Experiment 3 reviewed the effects of social consequences and information about expected behavior. The results suggested that social consequences (feedback) but not information about expected behavior change altered response bias among observers. Lerman et al. (2010) showed the preliminary viability of using an SD analysis for evaluating variables that may influence observer accuracy.

Smith et al. (2013) performed a similar experiment but focused on the clarity of an operational definition. Two different operational definitions were used to describe the same behavior: one focused on the events typically leading to the behavior and one associated with the socially significant behavior. A gold standard, or point of comparison, was created to calculate the accuracy of participant's answers. The SD analysis suggests that the operational definition of the socially significant behavior resulted in more consistent results relative to the gold standard than the operational definition about the events prior to the behavior occurring.

The gold standard method involves comparing a new measure to a current one that has high technically adequacy (Mudford et al., 2014; Smith et al., 2013). The "gold standard" method has been implemented within other behavior analytic studies as well. Poling et al. (2011) trained African pouched rats to detect landmines. To determine the extent to which the scent controlled the rats responding to the odor of the TNT, humans went over the area again with metal detectors and found that the rats did not miss any landmines. In this particular case, the metal detectors were the “gold standard” of comparison. African pouched rats were also trained to assess and diagnose the presence of tuberculosis (TB; Poling et al., 2011). The most commonly used diagnostic tool to assess TB in the developing world is analyzing sputum (i.e., mucus and phlegm) under a microscope which makes TB relatively easy to see and count (Dye et al., 2005, as cited in Poling et al., 2011). To determine the rats’ sensitivity and specificity of detection, the rats’ findings or diagnoses were compared to positive results identified via microscopy, the current gold standard. Finally, the “gold standard” method was used when developing the Functional Analysis Screening Tool (FAST) by comparing the results of the FAST to a functional analysis (FA; Iwata et al., 2013).

Purpose of the Present Study

In the present study, a digital version of the PDC-HS (Digital PDC-HS) was administered using Qualtrics® to determine barriers to fluent task completion and indicated interventions. Digital administration allowed for anonymity of participant responses. The primary dependent variables were the duration of puzzle solving and participant responses on the PDC-HS. The independent variables were the domain manipulations as they relate to the PDC-HS, comprising a variety of imposed and removed barriers, such as omission of training, as well as provision of background information, a job aid, task purpose, and feedback, among others. The research question of interest was: Will a digital version of the PDC-HS accurately identify barriers to fluent task completion?

The primary aim of the present study was to use a complex behavior to validate the Digital PDC-HS. Construct validity is used to determine how well an assessment measures what it is intended to measure (Kazdin, 1999; Messick, 1995). However, to demonstrate validity, reliability had to be demonstrated first. Therefore, reliability was determined by examining the test-retest reliability of the Digital PDC-HS. Test-retest reliability examines if assessments given at different times to the same respondents will generate similar results (Floyd et al., 2005; McIntosh et al., 2008). Test-retest reliability can be examined by calculating a coefficient; the closer the coefficient is to one, the higher test-retest reliability it has. To calculate the coefficient, a correlation is taken between the scores of the first test result compared to the second test result (McIntosh et al., 2008). However, in the present study, test-retest reliability was done using visual analysis. Test results were plotted and the more data points that fell on the center line indicated a higher test-retest reliability.

The second aim was to examine the extent to which the domain manipulations affected responding on the traditional interview-style version of the PDC-HS (Traditional PDC-HS) and the Digital PDC-HS. An SD analysis was conducted to quantify the differences, and assess the sensitivity, specificity, PPV, and NPV, of both the Digital and Traditional PDC-HS. Examining correlations between both administrations of the PDC-HS for the Traditional and the Digital PDC-HS groups, as well as a control group that experienced only baseline conditions (Baseline Control group in the first experiment), allowed for visual comparison of the two administrations between groups. The Traditional and Digital PDC-HS were compared using the "gold standard" method. Hypothetically, if the domain manipulations affected responding similarly on both assessments, their results would be the same.

Experiment 1

Method

Participants, Setting, and Materials

Participants.

The present translational study included twelve participants who were undergraduate students (11 females, 1 male) recruited from undergraduate courses in the Department of Applied Behavioral Science at the University of Kansas. Participants received three points or 0.5% of extra credit toward their overall grade. All individuals who expressed interest in the study qualified to participate, as none solved the sliding puzzle in under a minute three consecutive times during the initial baseline condition.

Per self-report, participants were undergraduate students between the ages of 19 and 23 (*M* = 20), with grade point averages (GPA) ranging from 2.80 to 3.98 (*M* = 3.46). Freshman, Sophomores, Juniors, and Seniors, and seven majors, were represented in the present study. The investigator read potential participants an oral consent statement (see Appendix A) and they were given an opportunity to ask questions. After individuals agreed to participate in the study, they were read a script that reviewed the study procedures (see Appendix B

Study Procedures**.**

Setting.

The experiment was conducted virtually using Zoom© Video Communications, Inc. software (see http://www.zoom.us). Participants viewed the puzzle on their personal computer in the browser of their choice. In the baseline condition, the browser tab containing the puzzle was the only tab visible to participants. In the remaining two conditions, more than one browser tab containing the puzzle may have been visible, as different browser tabs were used to present the written instructions and the background information.

Materials.

Sliding 3x3 puzzles were created at a commercial website (https://www.puzzel.org/en). The solution was the same for each puzzle (a completed picture) however, the number of moves required to solve the puzzle varied based on the puzzle's starting configuration (or the position of each of the 8 tiles when the puzzle is initially presented). Sliding 3x3 puzzles have 181,440 solvable tile configurations that vary by difficulty (Reinefeld, 1993). For all puzzles used in the present study, the timer and number of moves settings were turned off to prevent participants from seeing a real-time display of their puzzle solving duration and number of moves made. Additionally, the highest difficulty setting was enabled on all puzzles. Per the website, the larger number of shuffles beforehand, the more difficult the puzzle was to solve (Wustenraad, n.d.). Therefore, starting configurations that were presented to participants had a higher frequency of minimum moves relative to easier starting configurations. A Uniform Resource Locater (URL) for each puzzle trial was shared using the Zoom chat function. The puzzle, which was identical during all experimental conditions, was a picture of grass shown in Figure 1. During behavioral skills training (BST), participants rehearsed using a puzzle depicting a dog (see Figure 2). Two different puzzles were used to prevent maturation. Depending on the condition, participants were provided with written instructions that described how to solve the puzzle (see Appendix C). The written instructions included the written steps to solving the puzzle and diagrams displaying the appropriate placements of the numbered tiles at each step.

The Digital PDC-HS was administered and completed on participants' personal computers in the browser of their choice using Qualtrics® Research (http://www.qualtrics.com;

see Appendix D). Participants were provided with a URL for each administration of the Digital PDC-HS using the Zoom chat function. When the Traditional PDC-HS was administered, participants were asked the same questions face-to-face through Zoom and their answers were screen recorded for scoring purposes. Both versions of the PDC-HS were reworded to address performance as it relates to solving the puzzle (e.g., Are materials that are required for solving the puzzle readily available?). Also, some questions were tailored to the position held by the interviewee. For example, question number three in the training domain asks, “Can the employee state the purpose of the task?” (Carr & Wilder, 2016), however, the participant may not have been aware that they could or could not accurately state the purpose of the task so their answer of “yes” or “no” may not be accurate. Therefore, the question was rephrased into directions of “State the purpose of the task” and the participant response was rated by the investigator.

Dependent Variables and Data Collection Procedures

Primary Dependent Variables

Sessions were scheduled for 90 minutes and all data were collected within a single session. The primary dependent variables included the duration of puzzle solving and PDC-HS responses. The duration of puzzle solving was collected as the time elapsed (in seconds) between the onset and offset of the trial. Onset was defined as the investigator saying “start” and offset was defined as the participant solving the puzzle, the participant ending the trial, or the trial “timing out” after 120 s. The PDC-HS was administered twice: once before the domain manipulation condition and once after. An SD analysis was completed on the second administration of the PDC-HS to quantify the sensitivity, specificity, PPV, and NPV of both the Digital and Traditional versions of the PDC-HS.

To calculate the measures, responses were scored as true positives, false positives, true negatives, and false negatives. Results were determined by reviewing the PDC-HS results, participant performance and the contingencies they experienced. For example, a response would have been scored as a true positive if participant's response indicated that they received training and they had experienced trained. Scoring guidelines for each question are in Table 1.

Secondary Dependent Variables

The duration of BST was measured from the permanent product of the screen recorded session. BST duration was calculated from the onset of BST to the offset of BST, with onset defined as pressing play on the video that was shown to participants and the offset defined as the 100% completion of a puzzle in under a minute after all BST steps were mastered.

Collateral Measure

The minimum number of moves required to solve the puzzle was examined to ensure the starting configurations did not vary substantially within and across participants and groups. This analysis was completed because only the puzzle difficulty level, which constrained the range of starting configurations, could be controlled by the experimenter. The minimum number of moves required to solve the puzzle was determined by recording the starting configuration of all tiles at the onset of the trial and inputting it into a website that output the minimum number of moves required to solve the puzzle (https://gamingph.com/blackdesertmobile/puzzle/puzzle.html).

Additionally, data were collected on the degree to which questions on the Traditional PDC-HS were read verbatim during the interview (see Table 2). A verbal score was calculated on question delivery on the PDC-HS. Scores were calculated by dividing the number of questions asked verbatim by the total number of questions. Formulas were the same for rephrased questions, unprompted information, and clarified questions. However, dividend was the total number of rephrased questions, unprompted information, or clarified questions.

Interobserver Agreement

Secondary independent observers, trained to mastery, collected data on the following measures: the duration of puzzle solving, the minimum number of moves required to solve the puzzle, the verbal scores (verbatim, rephrase, unprompted, wait, and clarification) (see Table 2) and the SD scores (true positives, true negatives, false positives, and false negatives). Training consisted of BST and continued until the secondary observer and experimenter agreed on results obtained across three consecutive scoring opportunities. All data were collected using the permanent product of the screen recorded sessions or PDC-HS responses. If the number of trials within the condition was not divisible by 3, then the percentage was rounded up (e.g., IOA was calculated for 2 out of 4 trials). The puzzle trials for which Interobserver agreement (IOA) was scored were randomly selected using a Google LLC© number generator.

Reed and Azulay's (2011) IOA calculator was used to calculate all IOA measures. The appropriate IOA equation was chosen based on the dependent variable. Each IOA equation had a different page within the spreadsheet. Scores calculated by the primary observer were put in the first column in the corresponding spreadsheet and scores calculated by the secondary observer were put in the second column in the corresponding spreadsheet. The results were automatically calculated due to the equations imbedded into the IOA calculator.

IOA was calculated for the duration of puzzle solving for at least 33% of puzzle trials in each condition for each participant (Kratochwill et al., 2010). IOA on the duration of puzzle solving was calculated by dividing the smaller number by the larger number and multiplying by

100. IOA for the duration of puzzle solving for participants in the Traditional PDC-HS, Digital PDC-HS, and Baseline Control groups was 96%, 98%, and 99%, respectively.

IOA for the duration of BST was calculated for all participants. IOA on the duration of BST was calculated by dividing the shorter duration by the longer duration and multiplying by 100. IOA for duration of BST for the Traditional and Digital PDC-HS groups was 99%.

IOA was also calculated on the SD analysis and BST for 100% of the second administrations of both the Traditional and Digital PDC-HS. The SD scores were calculated by dividing the number of items agreed upon by the total number of items, multiplied by 100 (Bijou et al., 1968). IOA for the SD scores for the Traditional PDC-HS group ranged from 93% to 96% (*M* = 95%), and for the Digital PDC-HS group ranged from 94% to 98% (*M* = 95%).

IOA was calculated for the minimum number of moves for at least 33% of puzzle trials in each condition for each participant (Kratochwill et al., 2010). IOA was calculated by dividing the number of agreements by the number of agreements plus disagreements, and multiplying by 100 (Bijou et al., 1968). IOA for the minimum number of moves for participants in the Traditional PDC-HS, Digital PDC-HS, and Baseline Control groups was 100%, 100%, and 94%, respectively.

Finally, IOA was calculated on the verbal scores for 100% of the second administrations of the Traditional PDC-HS. IOA was calculated by dividing the number of agreements by the number of agreements plus disagreements, and multiplying by 100 (Bijou et al., 1968). IOA for the verbal scores for participants in the Traditional PDC-HS ranged from 85% to 95% (*M* = 90%).

Procedural Fidelity

Procedural fidelity was assessed for a total of at least 50% of the trials per condition, per participant, per experimental group. Trials were randomly selected using a Google LLC© number generator.

Secondary observers, trained to mastery using BST, used a checklist to score procedural fidelity, ensuring that puzzle trials were conducted as designed (see Table 3). Training consisted of BST and continued until the secondary observer and experimenter agreed on results obtained across three consecutive scoring opportunities. Procedural fidelity for the Traditional PDC-HS, Digital PDC-HS, and Baseline Control groups was 99% per group. A secondary observer collected data on 100% of intervention sessions to ensure that every participant received BST (Merritt, 2017). Procedural fidelity was 100%.

Secondary observers, trained to mastery, also reviewed 100% of the Digital PDC-HS and Traditional PDC-HS administrations to verify that all questions were asked and that the participants taking the Digital PDC-HS had at least five minutes to complete the Digital PDC-HS independently. The secondary observer reviewed anonymous Digital PDC-HS results (permanent products) and recorded whether each question was asked during the interview. Procedural fidelity for both the Traditional and the Digital PDC-HS was 100%.

Experimental Design

A nonconcurrent multiple baseline design across participants (Watson & Workman, 1981) was employed, with participants randomly assigned the number of trials conducted during the baseline condition (i.e., 3, 4, 5, or 6) using a Google LLC© number generator. When a participant entered the study, the number of baseline trials was randomly generated and assigned and was then removed from the possible number of baseline trials available for future participants (Watson & Workman, 1981).

Participants in the Baseline Control group experienced baseline contingencies throughout all three conditions. Participants in the Traditional and Digital PDC-HS groups were exposed to three experimental conditions: 1) baseline, 2) domain manipulation, and 3) barrier removal.

Procedures

Baseline

During the baseline condition, participants were read a script and asked to solve the puzzle (see Appendix E). Supplemental materials (i.e., written instructions), training, or background information were not provided. The investigator told participants that the trial would end after a pre-determined amount of time, when they solved a puzzle, or if they chose to end the trial. The investigator showed participants how to move the tiles via demonstration and told them to put the tiles in numerical order. The investigator demonstrated that tiles could be moved orthogonally as long as the tile was next to an empty square. Participants moved tiles themselves to demonstrate their understanding of how to move tiles. The timer began when participants were vocally prompted to start. Participants had two minutes to solve the puzzle. If they did not solve the puzzle within that time, they “timed-out” at 120 s. The baseline condition consisted of a predetermined number of baseline trials randomly assigned to each participant, however, when baseline data were variable steady state logic prevailed (Sidman, 1960).

First PDC-HS Administration

After participants completed the baseline condition, another script was read aloud (see Appendix F) where they were asked to complete the PDC-HS (see Appendix D). Depending on the group assignment, a Digital or Traditional PDC-HS was administered. Given that the Traditional PDC-HS had been validated through prior research by applying non-indicated and indicated interventions (Bowe & Sellers, 2018; Carr et al., 2013; Ditzian et al., 2015; Wilder et al., 2018), the results of the digital version of the PDC-HS completed by participants in both the Digital PDC-HS and Baseline Control groups were examined in relation to the results of the Traditional PDC-HS.

When participants were presented with the Digital PDC-HS, they were told that the investigator was going to leave the room and they had five minutes to take the survey within a browser on their computer. They were also told the results were anonymous (which they were). The investigator kept the Zoom camera on and left the room to mimic real-world applications of a Digital PDC-HS and to prevent clarifying questions from being asked. While administering the Traditional PDC-HS the investigator used natural wording and answered clarifying and follow up questions asked by participants.

Domain Manipulations

All participants within the Digital and Traditional PDC-HS groups experienced the same domain manipulations. Background information was read aloud and textually presented to participants (see Appendix G). During the baseline and domain manipulation conditions, the participant's environment was arranged to include barriers that would prevent participants from solving the puzzle fluently (i.e., omission of training). However, during the domain manipulation condition, manipulations representative of barriers in other PDC-HS domains were simultaneously removed to help participants solve the puzzle fluently (e.g., quiet work environment, written instructions, omission of competing tasks) and to decrease the likelihood domains other than *Training* would be endorsed by participants' responses on the second administration of the PDC-HS (see Appendix H). The domain manipulation condition continued until at least three data points were stable (Sidman, 1960). Specific domain manipulations are listed in Appendix H.

Training

Training was omitted to increase the likelihood that participants would identify lack of training as the primary barrier to puzzle completion on the PDC-HS. Therefore, participants were not trained on how to solve the sliding puzzle quickly. To be eligible for the study, participants could not already solve the puzzle three times consecutively in under a minute. The criterion was selected to mimic true work settings where an employee may demonstrate inconsistent task completion. Finally, participants were not told that they had to solve the puzzle in under a minute to decrease the likelihood they would state the target task on question two in the training domain on the PDC-HS.

Task Clarification and Prompting

Variables aligned with the *Task Clarification and Prompting* domain were manipulated to decrease the likelihood it would be identified as the primary barrier to task completion. Therefore, in the background information provided, participants were informed that the purpose of the task was to solve the puzzle to win the competition. Participants were provided with a job aid to help them complete the puzzle (i.e., written instructions) and verbally prompted to begin the task (see Appendix C).

Resources, Materials, and Processes

The *Resources, Materials, and Processes* domain was not intended to be identified as a barrier to task completion. Therefore, the background information stated that everyone in their "office" had been trained to complete the puzzle except for them. They were provided with materials that were considered by the investigator to be readily available, well organized, and optimally designed (see Appendices C and E). Finally, participants did not have to complete any competing tasks prior to solving the puzzle.

Performance Consequences, Effort, and Competition

The final domain, *Performance Consequences, Effort, and Competition* was not intended to be identified as a barrier to completion, therefore, vocal feedback was provided: “you solved the puzzle in X minutes.” Feedback was personalized to the amount of time (X) it took participants to solve the puzzle. If the puzzle was solved in under a minute, the investigator told participants “You scored a point in the puzzle competition.” If participants either timed out or completed the task in more than a minute, the investigator told participants that they did not complete the task in the given amount of time and therefore did not earn a point in the puzzle competition. Finally, no competing tasks were presented.

Second PDC-HS Administration

The PDC-HS was administered again after the domain manipulation condition (see Appendix D). If participants completed the Digital PDC-HS during the first administration, they completed the Digital PDC-HS during the second administration as well; if participants completed the Traditional PDC-HS during the first administration, they completed the Traditional PDC-HS during the second administration as well. The investigator told the participants they could refer back to the background information at any point while they were completing the PDC-HS.

Behavioral Skills Training

After the PDC-HS was administered for the second time, the domain manipulation or barrier was removed to verify that it was the manipulated domain that was the barrier to fluent task completion. Thus, for the participants in the present study, barrier removal involved providing participants with BST on how to employ an algorithm to aid puzzle solving.

To complete the first step, a participant had to locate tiles one, two, and three and place them in the top row horizontally in numerical order. To complete the second step, participants had to locate tiles four and seven and place them vertically in first column, or the column under tile one in numerical order. Finally, participants had to rotate tiles five, six, and eight until they were in the correct ending locations. All of these steps were completed in isolation meaning that once the previous steps were completed, the corresponding tiles were not moved to complete the succeeding step.

The training consisted of participants watching a one-and-a-half-minute video (How Kids Make Money, 2018) and receiving BST to learn the algorithm to solve the puzzle (see Appendix I). Participants were given an alternate puzzle to practice on to prevent testing effects (see Appendix D; Kazdin, 1999). Participants demonstrated each step twice while the investigator said the steps as they were completed (i.e., “First, I am going to place tile one in the top left hand corner”). After two demonstrations, they were also provided with steps to help them solve “tricky combinations” or configurations of tiles that were difficult to resolve. To train participants how to solve difficult configurations, the puzzle was purposefully put into a “tricky combination” and the investigator modeled how to solve it. After the tricky combination and steps had been demonstrated, participants rehearsed them until they completed the steps three times consecutively entirely independently. Participants were required to complete the entire puzzle three times and then once in under a minute before returning to the original puzzle to meet the mastery criterion. Once this criterion was met, they had completed BST. Any puzzle-related questions asked by participants were answered during BST.

Barrier Removal

Following BST, participants were asked to again solve the original puzzle. Trials continued until criterion was met and three data points were stable. Stability was determined by visual analysis and using steady state logic (Sidman, 1960).

Debriefing

Following the barrier removal condition, participants were read a brief statement (see Appendix J).

Data Analysis

Analysis of Verbal Scores

As natural language was used when administering the Traditional PDC-HS (in an effort to mimic real-world administrations), a post-hoc analysis was conducted to determine if the verbal behavior of question delivery affected participant responding. The extent to which participant responding was controlled by the verbatim verbal behavior of the question was calculated by adding the number of true positives and true negatives when questions were asked verbatim divided by the total number of questions asked verbatim. The degree to which participant responding was affected by the rephrased verbal behavior of question delivery was calculated by adding the total number of true positives and true negatives when questions were rephrased divided by the total number of rephrased questions.

Signal Detection

An SD analysis was performed for two purposes: 1) to better understand the stimulus control of participant responding and 2) measure the technical adequacy of the PDC-HS. For an assessment to have high technical adequacy, it must demonstrate high reliability and high validity (Gresham, 2003). Participant responses were divided into four categories to examine the stimulus control of participant responding: 1) hit or true positive; 2) false alarm or false positive; 3) correct rejection or true negative, and 4) miss or false negative (Treat & Viken, 2012). The goal of the PDC-HS is to determine the percentage of “No” responses in domains for the purpose of identifying performance barriers. Therefore, when a participant accurately identified an element of task completion missing, it was scored as a “true negative.” A true negative was defined as a participant accurately identifying a barrier (e.g., participants indicating that they had not received training when they truly had not). A true positive was defined as a participant accurately identifying that an element of task completion had taken place (e.g., participants indicating that they had solved the puzzle when they had). A false negative was defined as the participant stating that there was a barrier, when there wasn’t one (e.g., participants indicating that they had not been told to solve the puzzle even though they had). A false positive was defined as a participant neglecting to identify a barrier (e.g., participants indicating that they had received training although they had not). For example, question three under *Performance Consequences, Effort, and Competition* asks the employee if they ever see the effects of accurate task completion. If the participant reported, “Yes” and they had never solved the puzzle, the answer was scored as a “false positive” when interpreting the results of the PDC-HS.

Answers were scored on all questions and follow up questions. For example, question two in the *Resources, Materials, and Processes* domain asked participants to list the materials for the task. An answer was scored as a true positive if a material was identified correctly (i.e., puzzle or written instructions) and a false negative if any other material was listed inaccurately. Furthermore, some questions were scored based on the preceding answer. For example, questions three and four in *Resources, Materials, and Processes* asks if the materials were well designed and well organized, respectively. If the participant inaccurately identified the materials on the preceding question, the answer was automatically marked as a false positive or false negative depending on their response. The rational for this was that the participant was answering questions four and five in relation to the incorrect materials, and therefore, their answer was not reflecting the question.

Sensitivity, specificity, PPV, and NPV was used to quantify and measure the technical adequacy of the instrument. To determine the sensitivity and specificity of both the Digital and Traditional PDC-HS, an SD analysis was performed (Treat & Viken, 2012). Sensitivity was calculated by dividing the number of hits by the sum of true positives and false negatives (Treat & Viken, 2012). Specificity was calculated by dividing the number of correct rejections by the sum of true negatives and false positives (Treat & Viken, 2012). PPV was calculated by the number of true positives divided by the sum of true positives and false positives (Treat & Viken, 2012). Similarly, NPV was calculated by the number of true negatives divided by the sum of true negatives and false negatives (Treat & Viken, 2012). These four measures were examined at the group level.

Results

Sliding Puzzle Results

Figure 3 shows the duration of puzzle solving for participants in all three groups across all three conditions in Experiment 1. Scaled to the *x*-axis are conditions; scaled to the *y*-axis is median duration in seconds. The gray dotted line represents the maximum duration of a trial, or 120 s, and median duration of puzzle solving is denoted by the horizontal black line in each data cluster. The black data points represent participants in the Traditional PDC-HS group, the gray data points represent participants in the Baseline Control group, and the white data points represent participants in the Digital PDC-HS group.

In the baseline condition, the median duration of puzzle solving was above 100 s for 3 out of 4 participants. One participant from each group had median durations that were below 100 s. In the domain manipulation condition, the median duration of puzzle solving was similar to the median duration of puzzle solving evidenced during the baseline condition for all groups, suggesting that the domain manipulations provided (i.e., the non-indicated interventions) did not meaningfully improve puzzle solving for participants in the Traditional and Digital PDC-HS groups. In the barrier removal condition, 3 out of 4 participants in the Traditional and Digital PDC-HS groups met criterion. One participant from the Traditional PDC-HS and Digital PDC-HS groups, and all participants in the Baseline Control group, did not meet criterion. Taken together, these results suggest that BST was effective at increasing fluent task completion for 6 out of 8 participants.

Figures 4, 5, and 6 display the duration of puzzle solving for all individual participants, across all trials of all conditions, in the Traditional PDC-HS, Digital PDC-HS, and Baseline Control groups, respectively (see Appendix P).

Traditional PDC-HS Group

Figure 4 shows the duration of puzzle solving across participants for the experimental group administered the Traditional PDC-HS. During the baseline condition, Participant 32's duration of puzzle solving was stable at 120 s, indicating they never solved the puzzle. During the domain manipulation condition, the duration of puzzle solving remained stable at 120 s across all three trials. BST lasted 47 min and 17 s, therefore, the participant only had sufficient time left in the experimental session for one trial in the barrier removal condition. During the barrier removal condition, the duration of puzzle solving remained at 120 s.

During the baseline condition, Participant 43's data were variable, with the duration of puzzle solving ranging from 88 s to 120 s (*Mdn* = 107 s). During the domain manipulation condition, the overall level of the duration of puzzle solving increased slightly, ranging from 102 s to 120 s (*Mdn* = 110 s), with an upward trend evident. During the barrier removal condition following BST (which lasted 15 min and 59 s), the participant's duration of puzzle solving immediately decreased to a range of 18 s to 30 s (*Mdn* = 22 s).

During the baseline condition, Participant 36's duration of puzzle solving remained stable at 120 s across all nine baseline trials except for trial 6, thus the duration of puzzle solving ranged from 57 s to 120 s (*Mdn* = 120 s). During the domain manipulation condition, a decreasing trend was evident in the duration of puzzle solving, which ranged from 81 s to 120 s (*Mdn* = 108.5 s). During the barrier removal condition following BST (which lasted for 22 min and 0 s), the duration of puzzle solving ranged from 16 s to 95 s (*Mdn* = 36 s) and, with the exception of one outlier, evidenced a decreasing trend and lower overall level. Two data points overlap between the domain manipulation and barrier removal conditions, at 81 s and 97 s, respectively.

During the baseline condition, Participant 21's duration of puzzle solving was variable, ranging from 37 s to 120 s (*Mdn* = 71.5 s). During the domain manipulation condition, the duration of puzzle solving ranged from 18 s to 120 s (*Mdn* = 72 s). During the barrier removal condition following BST (which lasted for 14 min and 53 s), the duration of puzzle solving immediately decreased to a low stable level, ranging from 21 s to 27 s (*Mdn* = 25 s).

Baseline Control

Figure 5 shows the duration of puzzle solving across participants in the Baseline Control group. During the baseline condition, Participant 15's duration of puzzle solving remained at 120 s. During the second baseline condition, the duration of puzzle solving remained stable at 120 s except for Trial 5, thus the duration of puzzle solving ranged from 103 s to 120 s (*Mdn* = 120 s). During the final baseline condition, the duration of puzzle solving remained stable at 120 s except for Trial 7, thus the duration of puzzle solving ranged from 106 s to 120 s (*Mdn* = 120 s).

Participant 22 and Participant 20's duration of puzzle solving remained stable at 120 s across all trials in all three baseline conditions.

During the first baseline condition, Participant 6's data were variable. The duration of puzzle solving ranged from 20 s to 120 s (*Mdn* = 92 s). During the second baseline condition, the overall level of the duration of puzzle solving increased slightly, ranging from 36 s to 120 s (*Mdn* = 69 s), with an upward trend evident. During the third baseline condition, the duration of puzzle solving ranged from 36 s to 120 s (*Mdn* = 120 s).

Digital PDC-HS

Figure 6 shows the duration of puzzle solving across all participants in the Digital PDC-HS group. During the baseline condition, Participant 42's duration of puzzle solving remained stable at 120 s. During the domain manipulation condition, the duration of puzzle solving remained stable at 120 s. During the barrier removal condition following BST (which lasted for 36 min and 12 s), the duration of puzzle solving was 79 s. Due to the duration of BST, the participant only had sufficient time left in the experimental session for one trial in the barrier removal condition.

During the baseline condition, Participant 19's duration of puzzle solving remained stable at 120 s. During the domain manipulation condition, the duration of puzzle solving remained stable at 120 s. During the barrier removal condition following BST (which lasted 20 min and 54 s), the duration of puzzle solving immediately decreased to a range of 30 s to 55 s (*Mdn* = 43 s).

During the baseline and domain manipulation conditions, Participant 37's duration of puzzle solving remained stable at 120 s. During the barrier removal condition following BST (which lasted 13 min and 55 s), the duration of puzzle solving decreased immediately, ranging from 14 s to 29 s (*Mdn* = 18 s).

In the baseline condition, Participant 31's data were variable, with the duration of puzzle solving ranging from 38 s to 120 s (*Mdn* = 120 s). During the domain manipulation condition, data remained variable, with the duration of puzzle solving ranging from 28 s to 120 s (*Mdn* = 76 s). During the barrier removal condition following BST (which lasted 19 min and 47 s), data stabilized immediately, ranging from 15 s to 50 s (*Mdn* = 28 s).

Frequency Distribution of Minimum Moves Required

Figure 7 shows a frequency distribution of the minimum number of moves based on the starting configuration. The frequency distribution ranges from 10 to 28 indicating that starting configurations required between 10 and 28 minimum moves. The black histograms represent the baseline condition, the gray histograms represent the domain manipulation condition, and the white histograms represent the barrier removal condition. While some starting configurations required between 10 and 17 moves, most starting configurations required between 18 and 28 moves.

PDC-HS Results

Figure 8 displays the results of the PDC-HS averaged across all participant. The *x*-axis depicts the PDC-HS domains and the *y*-axis represents the percentage of questions scored “No” from each domain. The black histograms represent the Traditional PDC-HS group; the gray histograms represent the Baseline Control group, and the white histograms represent the Digital PDC-HS group.

Traditional PDC-HS Group

The results of the first administration of the Traditional PDC-HS indicated all domains of the PDC-HS were endorsed to varying degrees (see Figure 8). *Training* was endorsed the most, at an average of 68%; *Task Clarification and Prompting* was endorsed at an average of 65%; *Resources, Materials, and Processes* was endorsed the least, at an average of 28%; and *Performance Consequences, Effort, and Competition* was endorsed at an average of 45%.

The results of the second administration of the Traditional PDC-HS indicated endorsement of *Training* decreased from an average of 68% to 62%, yet *Training* was still endorsed the most; endorsement of *Task Clarification and Prompting* decreased from an average of 65% to 25%, thus *Task Clarification and Prompting* was endorsed the least; endorsement of *Resources, Materials, and Processes* increased from an average of 28% to 43%; and endorsement of *Performance Consequences, Effort, and Competition* decreased from an average of 45% to 30%.

Baseline Control Group

The results of the first administration of the Digital PDC-HS to participants in the

Baseline Control group indicated endorsement of all domains of the PDC-HS to varying degrees (see Figure 8). *Training* was endorsed the most, at an average of 87%; *Task Clarification and Prompting* was endorsed at an average of 60%; *Resources, Materials, and Processes* was endorsed the least, at an average at 40%; and *Performance Consequences, Effort, and Competition* was endorsed at an average of 75%.

The results of the second administration of the Digital PDC-HS to participants in the Baseline Control group indicated endorsement of *Training* decreased from an average of 87% to 75% and was tied for being endorsed the most; endorsement of *Task Clarification and Prompting* increased from an average of 60% to 75% and was tied for being endorsed the most; endorsement of *Resources, Materials, and Processes* decreased from an average of 40% to 28% and was endorsed the least; and endorsement of *Performance Consequences, Effort, and Competition* decreased from an average of 75% to 65%.

Digital PDC-HS Group

The results of the first administration of the Digital PDC-HS to participants in the Digital PDC-HS group indicated all domains of the PDC-HS were endorsed to varying degrees (see Figure 8). *Training* was endorsed at an average of 75%; *Task Clarification and Prompting* was endorsed at an average of 65%; *Resources, Materials, and Processes* was endorsed the most at an average of 77%; and *Performance Consequences, Effort, and Competition* endorsed the least at an average of 55%.

The results of the second administration of the Digital PDC-HS to participants in the Digital PDC-HS group indicated endorsement of *Training* decreased from an average endorsement of 75% to 68%, yet was still endorsed the most; endorsement of *Task Clarification and Prompting* decreased from an average of 65% to 35%, thus was endorsed the least;

endorsement of *Resources, Materials, and Processes* decreased from an average of 77% to 54%; and endorsement of *Performance Consequences, Effort, and Competition* stayed consistent at an average of 55%.

Correlations

The graphs in Figure 9 show the differences in the percentage of questions scored “No” between the 1st and 2nd administration of the PDC-HS. Scaled to the *x*-axis is the percentage of “No” responses of the 2nd Administration of the PDC-HS and scaled to the *y*-axis is the percentage of “No” responses of the 1st Administration of the PDC-HS. In the Traditional and Digital PDC-HS Group, participants experienced changing conditions. Different conditions were expected to yield different responses, therefore, data were expected to not fall on the center line. Furthermore, the baseline control graph was expected to have data on the center line as the conditions did not change.

The top graph depicts the correlation for the Traditional PDC-HS group (black symbols), the middle graph depicts the correlation for the Baseline Control group (gray symbols), and the bottom graph depicts the correlation for the Digital PDC-HS group (white symbols). Each data point represents the percentage of “No” responses for one participant on a specific domain of the PDC-HS. Circles represent the *Training* domain, diamonds represent the *Task Clarification and Prompting* domain, triangles represent the *Resources, Materials, and Processes* domain, and squares represent the *Performance Consequences, Effort, and Competition* domain.

For participants in the Baseline Control group, similar results were obtained between the first and second administrations of the PDC-HS, as evidenced by the majority of data points falling on the diagonal line. For participants in the Traditional PDC-HS group, with the exception of one data point representative of *Training* and one data point representative of

Performance Consequences, Effort, and Competition falling on the line, all other data appeared to be evenly distributed throughout the graph, suggesting little relation between the percentage of “No” responses on the first and second administration of the PDC-HS. For participants in the Digital PDC-HS group, with the exception of three data points representative of *Training* and one data point representative of *Performance Consequences, Effort, and Competition* falling on the line, most data appeared to be distributed across the top-left side of the graph, suggesting a higher percentage of “No” responses on the first administration relative to the second administration.

Individual Responses on the PDC-HS

The percentage of “No” responses to each question on the first and second administration of the PDC-HS was examined across all participants in the Traditional and Digital PDC-HS groups (see Figure 10). The table is divided by the first and second administration of the PDC-HS and the four domains. Green represents true positives, red represents true negatives, light green represents false positives, and maroon represents false negatives. As represented by the disproportionate amount of maroon, false negatives were higher in the Digital PDC-HS group than the Traditional PDC-HS group, which indicated that participants in the Digital PDC-HS group were identifying barriers that were intended to be removed by the background information provided. For example, on Question 1 in the Resources, Materials, and Processes domain, participants indicated that there were not enough employees trained to solve the puzzle when the background information stated otherwise.

Figure 11 graphically displays the number of participant responses in each of the four SD categories (i.e., true positive, false positive, true negative, and false negative).Scaled to the *x*-axis are the participants and scaled to the *y*-axis is the number of occurrences of each response. Each histogram represents one participant. Green represents true positives, light green represents false positives, red represents true negatives, and maroon represents false negatives (see Table 2). The legend applies to both graphs.

Traditional PDC-HS Group

On the first administration of the PDC-HS (see Figure 10), Participant 21 responded "No" to 2 out of 4 (50%) of the questions in the *Training* domain; 4 out of 5 (80%) of the questions in the *Task Clarification and Prompting* domain; 1 out of 5 (20%) of the questions in the *Resources, Materials, and Processes* domain; and 1 out of 5 (20%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 21 responded "No" to 8 out of 19 (42%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 10), Participant 21 responded "No" to 2 out of 4 (50%) of the questions in the *Training* domain; 0 out of 5 (0%) of the questions in the *Task Clarification and Prompting* domain; 0 out of 5 (0%) of the questions in the *Resources, Materials, and Processes* domain; and 0 out of 5 (0%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 21 responded "No" to 2 out of 19 (11%) of the questions on the second administration of the PDC-HS.

On the first administration of the PDC-HS (see Figure 10), Participant 43 responded "No" to 2 out of 4 (50%) of the questions in the *Training* domain; 2 of 5 (40%) of the questions in the *Task Clarification and Prompting* domain; 3 out of 6 (20%) of the questions in the *Resources, Materials, and Processes* domain; and 1 out of 5 (20%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 43 responded "No" to 8 out of 20 (40%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 10), Participant 43 responded

"No" to 3 out of 4 (75%) of the questions in the *Training* domain; 1 out of 5 (20%) of the questions in the *Task Clarification and Prompting* domain; 3 out of 6 (50%) of the questions in the *Resources, Materials, and Processes* domain; and 1 out of 5 (20%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 43 responded "No" to 8 out of 20 (40%) of the questions on the second administration of the PDC-HS.

On the first administration of the PDC-HS (see Figure 10), Participant 32 responded "No" to 4 out of 4 (100%) of the questions in the *Training* domain; 4 out of 5 (80%) of the questions in the *Task Clarification and Prompting* domain; 1 out of 5 (20%) of the questions in the *Resources, Materials, and Processes* domain; and 4 out of 5 (80%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 32 responded "No" to 13 out of 19 (68%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 10), Participant 32 responded "No" to 3 out of 4 (75%) of the questions in the *Training* domain; 2 out of 5 (40%) of the questions in the *Task Clarification and Prompting* domain; 4 out of 6 (66%) of the questions in the *Resources, Materials, and Processes* domain; and 3 out of 5 (60%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 32 responded "No" to 11 out of 20 (55%) of the questions on the second administration of the PDC-HS

On the first administration of the PDC-HS (see Figure 10), Participant 36 responded "No" to 3 out of 4 (75%) of the questions in the *Training* domain; 3 out of 5 (60%) of the questions in the *Task Clarification and Prompting* domain; 1 out of 5 (20%) of the questions in the *Resources, Materials, and Processes* domain; and 3 out of 5 (60%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 36 responded "No" to 10 out of 19 (53%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 10), Participant 36 responded "No" to 2 out of 4 (50%) of the questions in the *Training* domain; 2 out of 5 (40%) of the questions in the *Task Clarification and Prompting* domain; 2 out of 6 (33%) of the questions in the *Resources, Materials, and Processes* domain; and 2 out of 5 (40%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 36 responded "No" to 8 out of 20 (40%) of the questions on the second administration of the PDC-HS.

Digital PDC-HS Group

Participant responses were anonymous. Therefore, they do not have a participant number but a letter to differentiate between the responses. On the first administration of the PDC-HS (see Figure 10), Participant A responded "No" to 4 out of 4 (100%) of the questions in the *Training* domain; 4 out of 5 (80%) of the questions in the *Task Clarification and Prompting* domain; 5 out of 6 (83%) of the questions in the *Resources, Materials, and Processes* domain; and 5 out of 5 (100%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant A responded "No" to 18 out of 20 (90%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 10), Participant A responded "No" to 4 out of 4 (100%) of the questions in the *Training* domain; 2 out of 5 (40%) of the questions in the *Task Clarification and Prompting* domain; 4 out of 6 (66%) of the questions in the *Resources, Materials, and Processes* domain; and 4 out of 5 (80%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant A responded "No" to 13 out of 20 (65%) of the questions on the second administration of the PDC-HS.

On the first administration of the PDC-HS (see Figure 10), Participant B responded "No" to 2 out of 4 (50%) of the questions in the *Training* domain; 4 out of 5 (80%) of the questions in the *Task Clarification and Prompting* domain; 5 out of 6 (83%) of the questions in the *Resources, Materials, and Processes* domain; and 2 out of 5 (40%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant B responded "No" to 13 out of 20 (65%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 10), Participant B responded "No" to 1 out of 4 (25%) of the questions in the *Training* domain; 1 out of 5 (20%) of the questions in the *Task Clarification and Prompting* domain; 2 out of 6 (33%) of the questions in the *Resources, Materials, and Processes* domain; and 2 out of 5 (40%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant B responded "No" to 8 out of 20 (40%) of the questions on the second administration of the PDC-HS.

On the first administration of the PDC-HS (see Figure 10), Participant C responded "No" to 4 out of 4 (100%) of the questions in the *Training* domain; 3 out of 5 (60%) of the questions in the *Task Clarification and Prompting* domain; 1 out of 4 (25%) of the questions in the *Resources, Materials, and Processes* domain; and 2 out of 5 (40%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant C responded "No" to 10 out of 18 (56%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 10), Participant C responded "No" to 4 out of 4 (100%) of the questions in the *Training* domain; 2 out of 5 (40%) of the questions in the *Task Clarification and Prompting* domain; 4 out of 6 (66%) of the questions in the *Resources, Materials, and Processes* domain; and 4 out of 5 (80%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant C responded "No" to 14 out of 20 (70%) of the questions on the second administration of the PDC-HS.

On the first administration of the PDC-HS (see Figure 10), Participant D responded "No" to 2 out of 4 (50%) of the questions in the *Training* domain; 2 out of 5 (40%) of the questions in the *Task Clarification and Prompting* domain; 6 out of 6 (100%) of the questions in the *Resources, Materials, and Processes* domain; and 2 out of 5 (40%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant D responded "No" to 12 out of 20 (60%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 10), Participant D responded "No" to 2 out of 4 (50%) of the questions in the *Training* domain; 2 out of 5 (40%) of the questions in the *Task Clarification and Prompting* domain; 3 out of 6 (50%) of the questions in the *Resources, Materials, and Processes* domain; and 1 out of 5 (20%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant D responded "No" to 8 out of 20 (40%) of the questions on the second administration of the PDC-HS.

Verbal Scores

Figure 12 displays the verbal behavior of each question delivery stratified by the four SD categories for all participants in the Traditional PDC-HS group. Scaled to the *x*-axis is the verbal behavior of the question delivery (i.e., verbatim, rephrase, clarification, and unprompted; Table 2) and scaled to the *y*-axis is the number of occurrences of each response. Green represents true positives, light green represents false positives, red represents true negatives, and maroon represents false negatives.

Figure 12, shows that 33 out of 69 questions were asked verbatim. Of the 33 questions, 20 questions were scored as true positives, 1 question was scored as a false positive, 8 questions were scored as true negatives, and 4 questions were scored as false negatives. The results of the verbal behavior suggest that when the questions were asked verbatim, the participants identified 95% of components of task completion and 57% of barriers. Figure 12 also indicates that 35 out of 69 questions were rephrased when asked. Of the 35 questions rephrased, 15 were true positives, 9 were false positives, 5 were true negatives, and 6 were false negatives. The results of the rephrased verbal behavior indicate that the participants identified 65% of components of task completion and 56% of barriers. Participants only asked for clarifying information on three questions which resulted in two false negatives and one true positive. Four questions' verbal behavior were categorized as unprompted and resulted in two true negatives and two true positives.

Verbal scores were also analyzed by domain (see Figure 10). In the *Training* domain, 12 out of 16 (75%) of the questions were asked verbatim; 3 out of 16 (19%) of the questions were rephrased; 1 out of 16 (6%) of the questions were presented with unsolicited follow up information; and no clarifying information was provided.

In the *Task Clarification and Prompting* domain, 3 out of 20 (15%) of the questions were asked verbatim; 15 out of 20 (75%) of the questions were rephrased; 2 out of 20 (10%) of the questions were presented with unsolicited follow up information; and no clarifying information was provided.

In the *Resources, Materials, and Processes* domain, 8 out of 23 (35%) of the questions were asked verbatim; 12 out of 23 (52%) of the questions were rephrased; none of the questions were presented with unsolicited follow up information; and 2 out of 23 (9%) of the questions had clarifying information was provided.

In the *Performance Consequences, Effort, and Competition* domain, 12 out of 20 (60%) of the questions were asked verbatim; 6 out of 20 (30%) of the questions were rephrased; 1 out of 20 (5%) of the questions were presented with unsolicited follow up information; and 1 out of 20 (5%) of the questions had clarifying information provided.

Verbal scores were also calculated for each participant by summing the number of each score by the total number of questions asked on the PDC-HS (see Figure 10). For Participant 21, 8 out of 19 (42%) of the questions were asked verbatim; 11 out of 19 (58%) of the questions were rephrased; and no questions were presented with unsolicited follow up information or clarifying information.

For Participant 43, 10 out of 20 (50%) of the questions were asked verbatim; 7 out of 20 (35%) of the questions were rephrased; 2 out of 20 (10%) questions were presented with unsolicited follow up information; and 1 out of 20 (5%) questions were presented with clarifying information when the participant asked for it.

For Participant 32, 10 out of 20 (50%) of the questions were asked verbatim; 10 out of 20 (50%) of the questions were rephrased; and no questions were presented with unsolicited follow up information or clarifying information.

For Participant 36, 6 out of 20 (30%) of the questions were asked verbatim; 9 out of 20 (45%) of the questions were rephrased; 2 out of 20 (10%) questions were presented with unsolicited follow up information; and 2 out of 20 (10%) questions were presented with clarifying information when the participant asked for it.

Signal Detection

On the second administration of the PDC-HS, true positives, true negatives, false positives, and false negatives were scored for participants in both groups. The sensitivity, specificity, PPV, and NPV of both the Traditional and Digital PDC-HS were calculated (see Figure 20). For the Traditional PDC-HS, sensitivity ranged from 0.42 to 0.95 (M = 0.62), suggesting that the Traditional PDC-HS identified 62% of non-barriers to task completion. The specificity of the Traditional PDC-HS ranged from 0.68 to 0.84 (M = 0.77) suggesting that the

Traditional PDC-HS identified 77% of barriers to task completion. PPV ranged from 0.60 to 0.78 (*M* = 0.71) suggesting that 71% of identified non-barriers to task completion were not actually barriers. NPV ranged from 0.62 to 0.94 (*M* = 0.71) suggesting that 71% of identified barriers to task completion were true barriers.

For the Digital PDC-HS, sensitivity ranged from 0.31 to 0.64 (*M* = 0.51), suggesting that the Digital PDC-HS identified 51% of non-barriers to task completion. The specificity of the Digital PDC-HS ranged from 0.45 to 0.78 (*M* = 0.61), suggesting that the Digital PDC-HS identified 61% of barriers to task completion. PPV ranged from 0.33 to 0.74 (*M* = 0.58) suggesting that 58% of identified non-barriers to task completion are actually not barriers. NPV ranged from 0.44 to 0.72 (*M* = 0.56) suggesting that 56% of identified barriers to task completion are true barriers.

Individual SD Analyses

Traditional PDC-HS Group

Participant 32's graph (see Figure 11) shows that their responses consisted of 6 true positives, 2 false positives, 6 true negatives, and 6 false negatives. Participant 32's results indicate that the participant identified 66% of components of task completion and identified 50% of barriers.

Participant 43's graph (see Figure 11) shows that their responses consisted of 10 true positives, 2 false positives, 4 true negatives, and 6 false negatives. Participant 43's results indicate that the participant identified 83% of components of task completion and 40% of barriers.

Participant 36's graph (see Figure 11) shows that their responses consisted of 9 true positives, 3 false positives, 4 true negatives, and 4 false negatives. Participant 32's results indicate that the participant identified 75% of components of task completion and identified 50% of barriers.

Participant 21's graph (see Figure 11) shows that their responses consisted of 14 true positives, 3 false positives, 2 true negatives, and 0 false negatives. Participant 21's results indicate that the participant identified 82 % of components of task completion 100% of barriers.

Digital PDC-HS Group

Participant A's graph (see Figure 11) shows that their responses consisted of 6 true positives, 0 false positives, 5 true negatives, and 8 false negatives. Participant A's results indicate that the participant identified 100% of components and identified 38% of barriers.

Participant B's graph (see Figure 11) shows that their responses consisted of 12 true positives, 2 false positives, 1 true negative, and 4 false negatives. Participant B's results indicate that the participant identified 68% of components of task completion and identified 20% of barriers.

Participant C's graph (see Figure 11) shows that their responses consisted of 6 true positives, 0 false positives, 6 true negatives, and 8 false negatives. Participant C's results indicate that the participant identified 100% of components of task completion and identified 43% of barriers.

Participant D's graph (see Figure 11) shows that their responses consisted of 6 true positives, 6 false positives, 2 true negatives, and 5 false negatives. Participant D's results indicate that the participant identified 50% of components of task completion and identified 29% of barriers.

Discussion

After BST was implemented with participants in the Traditional and Digital PDC-HS groups, the duration of puzzle solving decreased and stabilized for 7 out of 8 participants, and 6 participants met mastery criterion (i.e., completing the puzzle in under a min across three consecutive trials). Participants in the Baseline Control group, who were not trained to solve the puzzle using BST, never met mastery criterion and their data either stabilized around 120 s or continued to be variable throughout all conditions, providing further evidence that lack of training was the primary barrier to fluent task completion.

Two out of 8 participants did not meet the fluent puzzle solving criterion despite BST taking place, which is one limitation of the present study. The null results could be contributed to different factors. The first possible explanation is time constraints. Participants only consented to 90 minutes of participation. Both participants who did not meet mastery criterion took substantially longer to complete BST than the other participants; the mean duration of BST for participants who met the fluent puzzle solving criterion was 17 min and 44 s as compared to the mean duration of BST for participants who did not meet the fluent puzzle solving criterion, which was 41 min and 44 s. Due to the long duration of BST, both participants only had enough time to complete one trial during the barrier removal condition, which prevented the investigator from collecting data on the duration of puzzle solving to stability. Another explanation for the null results could be that lack of training was not the only barrier participants experienced. While previous studies demonstrated undifferentiated responding and addressed only one barrier, it appears that the undifferentiated responding in this study may have suggested more than one barrier. Furthermore, the results could suggest that the procedure needs to continue to be developed until accurate responding occurs on the Traditional PDC-HS.

The results of the frequency distribution suggest that the range of the minimum number of moves required to complete the puzzle varied based on the starting configuration. However, despite the variability, the duration of puzzle solving remained consistent between participants, suggesting that starting configuration and minimum number of moves did not have an effect on the duration of puzzle solving.

The results of the second administration of the PDC-HS across both the Digital and the Traditional PDC-HS groups indicate *Training* is the primary barrier to fluent task completion. While the results of the second administration of the PDC-HS also showed somewhat undifferentiated responding across all other domains, the duration of puzzle solving did not decrease in the domain manipulation condition, suggesting that removal of barriers related to *Task Clarification and Prompting*, *Resources*, *Materials and Processes*, as well as *Performance Consequences, Effort, and Competition* functioned as a non-indicated intervention. The consistent and immediate decrease in the duration of puzzle solving after BST was implemented suggest that the functional barrier was indeed *Training* and the other barriers, despite the somewhat undifferentiated responding across domains on the PDC-HS, were not barriers.

A correlation was conducted between the percentage of “No” responses on the first and second administrations of the PDC-HS to examine the extent to which responding differed from the first administration to the second administration (see Figure 9). As expected, there appeared to be a strong correlation between the percentage of “No” responses across the two PDC-HS administrations for participants in the Baseline Control group who did not experience any differential environmental manipulations across conditions. Only low to moderate correlations appeared to be evident between the percentage of “No” responses on the two administrations of the PDC-HS for participants in the Traditional and Digital PDC-HS groups. However, only moderate correlations were expected because the *Training* domain was the only domain for which the percentage of “No” responses was anticipated to remain the same during the first and second administrations. Participants’ “No” responses in all other domains except *Training* were anticipated to have identified those domains as barriers on the first administration of the PDC-HS, but not on the second administration, which would result in the majority of the data points being plotted on the *y*-axis. Barriers in all domains that were anticipated to be identified on the first administration as barriers had not yet been removed to promote fluent puzzle solving. While lack of training was consistently identified as the primary barrier, endorsement of other domains indicated participants in both the Traditional and Digital PDC-HS groups identified additional barriers to puzzle completion.

The visual analysis of the correlation graphs suggest that training was consistently seen as a barrier, but other domains were as well. Therefore, the domain manipulations did not demonstrate control over participant responding. Furthermore, the visual display allowed for swift visual analysis and interpretation of the participant responses on the PDC-HS in relation to the domain manipulations.

A first step towards understanding the differences in responding on the Traditional and Digital PDC-HS involved reviewing the interviews given to participants in the Traditional PDC-HS group. The effects of the verbatim and rephrased verbal behaviors were examined to determine the effect verbal behavior had on participant responding. The results of the analyses of the verbal scores suggested that the verbal behavior of question delivery did not consistently impact participant responding on the second administration of the Traditional PDC-HS (see Figure 10).

To further explore possibilities for why domains other than the *Training* domain were endorsed, an SD analysis was performed to examine the stimulus control of participants' self-reported responses on the PDC-HS (i.e., true positives, false positives, true negatives, and false negatives). As represented by the disproportionate amount of maroon cells in Figure 10, false negatives were higher in the Digital PDC-HS group than the Traditional PDC-HS group, which indicated that participants in the Digital PDC-HS group were identifying barriers that were intended to be removed by the background information provided. For example, on Question 1 in the *Resources, Materials, and Processes* domain, participants indicated that there were not enough employees trained to solve the puzzle when the background information stated otherwise.

Additionally, the experimental manipulations were examined in relation to the PDC-HS results. Participants in both the Traditional and Digital PDC-HS groups tacted the written instructions as a form of training, which was scored as a false positive. Other additional information that had been provided to participants via the background information (such as the purpose of solving the puzzle, the number of trained employees, and materials) was also not always reflected in participants (from either of the groups) responses on the PDC-HS (see Appendix H). Contingency specifying stimuli, such as a discriminative stimulus, can have a function-altering effect on establishing operations that may result in rule-governed behavior (RGB; Schlinger & Blakely, 1987). RGB occurs when a behavior is maintained by antecedents as opposed to contingencies (Skinner, 1969). RGB has been shown to override experienced contingencies (Lippman & Meyer, 1967). Furthermore, Ayllon and Azrin (1964) showed that they were unable to gain schedule control until instructions were given about the desired response. Finally, studies have shown that instructions may facilitate the development of schedule control (Turner & Solomon, 1962). Therefore, the possibility that participants were not attending to the background information, provided in an effort to govern participant responding on the PDC-HS, was explored.

Sensitivity, specificity, positive predictive value, and negative predictive value were calculated in an effort to quantify overall responding on the Traditional and Digital PDC-HS (see Table 4). The results of the SD analysis suggest that not only was the stimulus control of participant responding not being maintained by the background information provided, experienced contingencies, and participant performance, but both the Traditional and Digital PDC-HS had low technical adequacy. The low technical adequacy of the Traditional PDC-HS is of higher priority than the Digital PDC-HS, as the Traditional PDC-HS should be the “gold standard” against which the Digital PDC-HS is compared.

Experiment 2

Method

The purpose of Experiment 2 was to examine the extent to which an animated video (instead of written and oral background information) and a checklist (instead of written instructions controlled participant responses during the second administration of the PDC-HS. Changing the delivery of the background information to an animated video was expected to change participants' self-reported responses on the second administration of the PDC-HS. Furthermore, changing the written instructions to a checklist was expected to decrease the likelihood of participants reporting they had not been trained. However, the checklist was substituted for the written instructions so that participants would still be able to identify it as a job aid and assess its organization and utility as a material.

Participants, Setting, and Materials

Three undergraduate females and one undergraduate male ranging in age from 19 to 21 years (M = 20) were recruited from applied behavior analysis classes. GPAs ranged from 3.40 to 3.94 (M = 3.68). Freshman, Sophomores, and Juniors, and two majors, were represented in the present study. The setting was identical to that described in Experiment 1.

An animated video was created that provided the same background information as the written instructions. The video had a voice over as well as background music (see Appendix K). The video can be found at (https://explee.com/video/0rpwd95; Figure 13). The video was watched twice – once after the first administration of the PDC-HS and again before the second administration of the PDC-HS. The animated video replaced the oral and written background information provided in Experiment 1. Participants were also provided with a checklist (see Appendix L) instead of written instructions to evaluate if they would tact the checklist as a job aid in the *Task, Clarification, and Prompting* domain but not as a form of training in the *Training* domain on the second administration of the PDC-HS.

The PDC-HS and puzzles were the same as those used in Experiment 1. However, during the last two conditions, participants could only have the puzzle and checklist open and visible in the browser of their choice, as no written background information was provided.

There were two groups of participants who experienced the same procedures and conditions (but with adapted domain manipulations, as discussed below) as their respective groups in Experiment 1 Traditional PDC-HS group and the Digital PDC-HS .

Dependent Variables and Data Collection Procedures

The dependent variables and data collection procedures were the same as in Experiment 1.

Interobserver Agreement

IOA calculations were the same as Experiment 1. One additional calculation was added for the use of the checklist which was calculated by Trial-by-Trial agreement. IOA for the duration of puzzle solving for participants in the Traditional PDC-HS and Digital PDC-HS groups were both 99%. IOA for BST was 99%. IOA for the minimum number of moves for participants in the Traditional PDC-HS and Digital PDC-HS was 88% and 100%, respectively. IOA on the verbal scores for participants in the Traditional PDC-HS group was 90%. IOA for the SD scores ranged from 93% to 96% (*M* = 95%) for the Traditional PDC-HS group and 93% to 97% (*M* = 95%) for the Digital PDC-HS. IOA for the amount of times the checklist was used was 100%.

Procedural Fidelity

Procedural fidelity calculations were the same as in Experiment 1. Procedural fidelity for the Traditional PDC-HS and Digital PDC-HS groups were calculated to be 99% and 98%, respectively.

Experimental Design

The experimental design was the same as in Experiment 1.

Procedures

All procedures were the same as in Experiment 1, except for those employed during the domain manipulation condition.

During the domain manipulation condition, participants in the Traditional and Digital PDC-HS groups experienced the same domain manipulations. Participants were provided with a checklist and told it was to help them track their accuracy of puzzle solving (see Appendix L). Participants also watched an animated video twice, once at the beginning and once at the end of the domain manipulation condition.

Two participants were administered a Traditional PDC-HS and the other two participants were administered a Digital PDC-HS, with administrations occurring after the baseline and domain manipulation conditions, as during Experiment 1.

Data Analysis

Analysis of Verbal Scores

Analysis of verbal scores were the same as in Experiment 1.

Signal Detection

The SD analysis was the same as in Experiment 1.

Results

Sliding Puzzle Results

Figure 14 shows the duration of puzzle solving for participants in both groups across all three conditions in Experiment 2. Graphing conventions remain the same as Experiment 1. In the baseline condition, the median duration of puzzle solving was above 110 s for all participants in both the Traditional and Digital PDC-HS groups. In the domain manipulation condition, the median duration of puzzle solving was similar to baseline for both participants in the Traditional PDC-HS group and 1 out of 2 participants in the Digital PDC-HS group; the remaining participant's median duration of puzzle solving decreased, however, did not meet criterion suggesting that the domain manipulations provided, or the non-indicated interventions, did not meaningfully improve puzzle solving for participants in the Traditional and Digital PDC-HS groups. In the barrier removal condition, all participants in the Traditional and Digital PDC-HS groups met criterion, suggesting that BST was effective at increasing fluent task completion for all participants.

Figures 15 and 16 display the duration of puzzle solving across participants in the experimental groups administered the Traditional PDC-HS and Digital PDC-HS, respectively.

Traditional PDC-HS Group

During the baseline condition, Participant 46's data were stable at 120 s, indicating the participant never solved the puzzle. During the domain manipulation condition, the duration of puzzle solving remained stable at 120 s. During the barrier removal condition following BST (which lasted 28 min and 42 s), the duration of puzzle solving immediately decreased, ranging from 21 s to 43 s (*Mdn* = 23 s). Participant 46 used the checklist to track progress from the baseline condition, 100% during the domain manipulation condition, and 0% on the barrier removal condition.

During the baseline condition, Participant 45's data were stable at 120 s, indicating they never solved the puzzle. During the domain manipulation condition, the duration of puzzle solving remained stable at 120 s. During the barrier removal condition following BST (which lasted 23 min and 1 s), the duration of puzzle solving immediately decreased, ranging from 21 s to 58 s (*Mdn* = 30.5 s). Participant 45 used the checklist to track progress from the baseline condition, 100% during the domain manipulation condition, and 100% on the barrier removal condition. Furthermore, when the checklist was originally presented, they filled in their results from the baseline condition. Finally, the checklist was not used during Trial 12, but was filled in when inputting data from Trial 13.

Digital PDC-HS Group

During the baseline condition, Participant 44's duration of puzzle solving remained stable at 120 s. During the domain manipulation condition, the duration of puzzle solving remained stable at 120s. During the barrier removal condition following BST (which lasted 17 min and 56 s), the duration of puzzle solving immediately decreased, ranging from 20 s to 51 s (*Mdn* = 31 s).

In the baseline condition, Participant 47's data were variable, with the duration of puzzle solving ranging from 108 s to 120 s (*Mdn* = 120 s). During the domain manipulation condition, data remained variable and the duration of puzzle solving ranged from 58 s to 120 s (*Mdn* = 120 s). During the barrier removal condition following BST (which lasted 22 m and 22 s), the duration of puzzle solving immediately decreased, ranging from 30 s to 64 s (*Mdn* = 42 s). Two data points overlap between the domain manipulation and barrier removal conditions, at 58 s and 64 s, respectively.

Due to the anonymous results, it is not clear which participant is associated with the responses of Participants E and F. However, of the two participants in the Digital PDC-HS group

(Participants 44 and 47), Participant 44 used the checklist to track progress from the baseline condition, 0% during the domain manipulation condition, and 0% on the barrier removal condition. Participant 47 did not track progress from the baseline condition, used the checklist for 100% of the domain manipulation condition, and 0% of the barrier removal condition.

Frequency Distribution of Minimum Moves Required

Figure 17 shows a frequency distribution of the minimum number of moves based on the starting configuration. The frequency distribution ranges from 14 to 26, indicating that the puzzles required between 14 and 26 minimum moves to solve. The black histograms represent the baseline condition, the white histograms represent the domain manipulation condition, and the gray histograms represent the barrier removal condition. While there were some starting configurations between 14 and 17, most starting configurations were between 18 and 26.

PDC-HS Results

Figure 19 displays the results of the PDC-HS averaged across both participants. The graphing conventions remain the same as Experiment 1. The results of the first administration of the Traditional PDC-HS indicated all domains of the PDC-HS were endorsed to varying degrees. *Training* was endorsed the most, at an average of 100%; *Task Clarification and Prompting* was endorsed the least, at an average of 60%; *Resources, Materials, and Processes* was endorsed at an average of 81%; and *Performance Consequences, Effort, and Competition* was endorsed the least at an average of 60%.

The results of the second administration of the Traditional PDC-HS indicated endorsement of *Training* remained consistent at an average of 100% and *Training* was still endorsed the most; endorsement of *Task Clarification and Prompting* decreased from an average of 60% to 30%; endorsement of *Resources, Materials, and Processes* decreased from an average of 81% to 0% (and *Resources, Materials, and Processes* was now endorsed the least); and endorsement of *Performance Consequences, Effort, and Competition* decreased from an average of 60% to 20%.

The results of the first administration of the Digital PDC-HS indicated all domains of the PDC-HS were endorsed to varying degrees. *Training* was endorsed the most, at an average of 100%; *Task Clarification and Prompting* was endorsed at an average of 70%; *Resources, Materials, and Processes* was endorsed at an average of 81%; and *Performance Consequences, Effort, and Competition* was endorsed the least, at an average of 60%.

The results of the second administration of the Digital PDC-HS indicated all domains decreased in the amount that they were endorsed. Endorsement of *Training* decreased from an average of 100% to 80%, yet *Training* was still endorsed the most; endorsement of *Task Clarification and Prompting* decreased from an average of 70% to 30% and *Task Clarification and Prompting* was endorsed the least; endorsement of *Resources, Materials, and Processes* decreased from an average of 81% to 63%, and endorsement of *Performance Consequences, Effort, and Competition* decreased from an average of 60% to 50%.

Finally, the use of the checklist varied between participants within and across the Traditional and Digital PDC-HS groups. Therefore, there does not appear to be a relation between use of checklist and the identification of it as a job aid or listed in materials on the PDC-HS.

Correlations

The graphs in Figure 19 show the differences in the percentage of questions scored “No” between the 1st and 2nd administration of the PDC-HS. The graphing conventions remain the same as Experiment 1.

The top graph depicts the correlation for the Traditional PDC-HS group (black symbols), and the bottom graph depicts the correlation for the Digital PDC-HS group (white symbols). Graphing conventions remain the same from Experiment 1. For participants in the Traditional PDC-HS group, with the exception of one data point representative of *Training*, all other data appeared to be either on the *y*-axis or above the diagonal line, suggesting a higher percentage of "No" responses on the first administration relative to the second administration.

For participants in the Digital PDC-HS group, with the exception of one data point representative of *Training* and one data point representative of *Performance Consequences, Effort, and Competition* falling on the line, most data appeared to be distributed across the top-left side of the graph, suggesting a higher percentage of "No" responses on the first administration relative to the second administration.

Individual Responses on the PDC-HS

The percentage of "No" responses to each question on the first and second administration of the PDC-HS were examined across all participants in the Traditional and Digital PDC-HS groups (see Figure 20). The graphing conventions remain the same from Experiment 1.

Traditional PDC-HS Group

On the first administration of the PDC-HS (see Figure 20), Participant 45 responded "No" to 4 out of 4 (100%) of the questions in the *Training* domain; 3 out of 5 (60%) of the questions in the *Task Clarification and Prompting* domain; 3 out of 5 (60%) of the questions in the *Resources, Materials, and Processes* domain; and 2 out of 5 (40%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 45 responded "No" to 12 out of 19 (63%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 20), Participant 45 responded

“No” to 4 out of 4 (100%) of the questions in the *Training* domain; 1 out of 5 (20%) of the questions in the *Task Clarification and Prompting* domain; 0 out of 5 (0%) of the questions in the *Resources, Materials, and Processes* domain; and 0 out of 5 (0%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 45 responded “No” to 5 out of 19 (26%) of the questions on the second administration of the PDC-HS.

On the first administration of the PDC-HS (see Figure 20), Participant 46 responded “No” to 4 out of 4 (100%) of the questions in the *Training* domain; 3 out of 5 (60%) of the questions in the *Task Clarification and Prompting* domain; 1 out of 5 (20%) of the questions in the *Resources, Materials, and Processes* domain; and 3 out of 5 (60%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 46 responded “No” to 11 out of 19 (58%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 20), Participant 46 responded “No” to 4 out of 4 (100%) of the questions in the *Training* domain; 2 out of 5 (40%) of the questions in the *Task Clarification and Prompting* domain; 0 out of 5 (0%) of the questions in the *Resources, Materials, and Processes* domain; and 2 out of 5 (40%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant 46 responded “No” to 8 out of 19 (42%) of the questions on the second administration of the PDC-HS.

Digital PDC-HS Group

On the first administration of the PDC-HS (see Figure 20), Participant E responded “No” to 4 out of 4 (100%) of the questions in the *Training* domain; 3 out of 5 (60%) of the questions in the *Task Clarification and Prompting* domain; 6 out of 6 (100%) of the questions in the *Resources, Materials, and Processes* domain; and 3 out of 5 (60%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant E responded

"No" to 16 out of 20 (80%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 20), Participant E responded "No" to 2 out of 4 (50%) of the questions in the *Training* domain; 1 out of 5 (20%) of the questions in the *Task Clarification and Prompting* domain; 4 out of 6 (66%) of the questions in the *Resources, Materials, and Processes* domain; and 3 out of 5 (60%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant E responded "No" to 10 out of 20 (50%) of the questions on the second administration of the PDC-HS.

On the first administration of the PDC-HS (see Figure 20), Participant F responded "No" to 4 out of 4 (100%) of the questions in the *Training* domain; 4 out of 5 (80%) of the questions in the *Task Clarification and Prompting* domain; 3 out of 5 (60%) of the questions in the *Resources, Materials, and Processes* domain; and 3 out of 5 (60%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant F responded "No" to 14 out of 19 (74%) of the questions on the first administration of the PDC-HS.

On the second administration of the PDC-HS (see Figure 20), Participant 46 responded "No" to 4 out of 4 (100%) of the questions in the *Training* domain; 1 out of 5 (20%) of the questions in the *Task Clarification and Prompting* domain; 4 out of 5 (80%) of the questions in the *Resources, Materials, and Processes* domain; and 2 out of 5 (40%) of the questions in the *Performance Consequences, Effort, and Competition* domain. Overall, Participant F responded "No" to 11 out of 19 (58%) of the questions on the second administration of the PDC-HS.

Verbal Scores

In Experiment 2, domain manipulations were adjusted with the intention to further control participant responding. Figure 22 graphically represents the verbal behavior for the Traditional PDC-HS group. Twenty-two out of 38 questions were asked verbatim. Of the 22 questions asked verbatim, 13 questions were scored as true positives, one question was scored as a false positive, eight questions were scored as true negatives, and zero questions were scored as a false negative. The results suggest that verbatim verbal behavior controlled participant responding as they identified 93% of components of task completion and 100% of barriers.

Fourteen out of 38 questions were rephrased when delivered to participants. Of the 14 questions rephrased, 6 questions were scored as true positives, 5 questions were scored as false positives, 2 questions were scored as true negatives, and 2 questions were scored as false negatives. The results suggest that rephrased verbal behavior controlled participant responding as they identified 70% of components of task completion and 50% of barriers.

Verbal scores were also analyzed by domain (see Figure 20). In the *Training* domain, 7 out of 8 (88%) of the questions were asked verbatim; 1 out of 8 (12%) of the questions were rephrased; no questions were presented with unsolicited follow up information; and no clarifying information was provided.

In the *Task Clarification and Prompting* domain, 2 out of 10 (20%) of the questions were asked verbatim; 7 out of 10 (70%) of the questions were rephrased; 1 out of 10 (10%) of the questions were presented with unsolicited follow up information; and no clarifying information was provided. In the *Resources, Materials, and Processes* domain, 5 out of 10 (50%) of the questions were asked verbatim; 5 out of 10 (50%) of the questions were rephrased; no questions were presented with unsolicited follow up information; and no clarifying information was provided. In the *Performance Consequences*, *Effort, and Competition* domain, 8 out of 10 (80%) of the questions were asked verbatim; 2 out of 10 (20%) of the questions were rephrased; no questions were presented with unsolicited follow up information; and no clarifying information was provided.

Verbal scores were also calculated for each participant by summing the number of each score by the total number of questions asked on the PDC-HS (see Figure 20). For Participant 45, 10 out of 19 (53%) of the questions were asked verbatim, 9 out of 19 (47%) of the questions were rephrased, and no questions were presented with unsolicited follow up information or clarifying information.

For Participant 46, 12 out of 19 (63%) of the questions were asked verbatim; 6 out of 19 (32%) of the questions were rephrased; 1 out of 19 (5%) questions were presented with unsolicited follow up information; and 0 out of 19 (0%) questions were presented with clarifying information when the participant asked for it.

Signal Detection

A SD analysis was conducted to determine the sensitivity, specificity, PPV, and NPV of both the Traditional and Digital PDC-HS (see Table 5). The sensitivity was consistent for both participants who were administered the Traditional PDC-HS at 0.88, suggesting that the Traditional PDC-HS identified 88% of non-barriers to task completion. The specificity of the Traditional PDC-HS ranged from 0.72 to 0.83 (*M* = 0.78) suggesting that the Traditional PDC-HS identified 78% of barriers to task completion. PPV ranged from 0.65 to 0.75 (*M* = 0.70), suggesting that 70% of identified non-barriers to task completion are actually not barriers. NPV ranged from 0.91 to 0.92 (*M* = 0.92) suggesting that 92% of identified barriers to task completion are true barriers.

For the Digital PDC-HS, sensitivity ranged from 0.50 to 0.60 (*M* = 0.55), suggesting that the Digital PDC-HS identified 55% of non-barriers to task completion. The specificity of the Digital PDC-HS ranged from 0.43 to 0.96 (*M* = 0.69), suggesting that the Digital PDC-HS identified 69% of barriers to task completion. PPV ranged from 0.65 to 0.92 (*M* = 0.79), suggesting that 79% of identified non-barriers to task completion are actually not barriers. NPV ranged from 0.29 to 0.74 (M = 0.51) suggesting that 51% of identified barriers to task completion are true barriers.

Individual SD Analyses

Traditional PDC-HS Group

Participant 46's graph (see Figure 22) shows that their responses consisted of 9 true positives, 2 false positives, 7 true negatives, and 1 false negative. Participant 46's results indicate that the participant identified 82% of components of task completion and identified 88% of barriers.

Participant 45's graph (see Figure 22) shows that their responses consisted of 10 true positives, 4 false positives, 4 true negatives, and 1 false negative. Participant 45's results indicate that the participant identified 71% of components of task completion and identified 80% of barriers.

Digital PDC-HS Group

Participant E's graph (see Figure 22) shows that their responses consisted of 10 true positives, 0 false positives, 2 true negatives, and 7 false negatives. Participant E's results indicate that the participant identified 100% of components of task completion and identified 31% of barriers.

Participant F's graph (see Figure 22) shows that their responses consisted of 6 true positives, 2 false positives, 6 true negatives, and 5 false negatives. Participant F's results indicate that the participant identified 75% of components of task completion and identified 55% of barriers.

Discussion

The purpose of Experiment 2 was to examine the effects of an animated video providing background information and replacing written instructions with a checklist on the control of participant responding on the Traditional and Digital versions of the PDC-HS. The results suggest that the animated video providing background information and checklist controlled participant responding more consistently than the domain manipulations they replaced in Experiment 1.

The domain manipulations appeared to exert some control over participant responding on the PDC-HS, as participants no longer indicated that they had received training, however the extent to which the checklist was considered a job aid varied across participants. In the Digital PDC-HS group, both Participants E and F indicated that a job aid was still available and labeled the puzzle and checklist as materials. In the Traditional PDC-HS group, Participant 46 indicated that a job aid was available and labeled the checklist as a material, but Participant 45 did not. Both Participants 45 and 46 also identified a variety of other materials, such as WiFi, computer, etc. Therefore, the domain manipulations (e.g., background information and checklist) controlled participant responding on the Digital PDC-HS but not the Traditional PDC-HS. That said, the small group size limits interpretation.

Similar to Experiment 1, results of the frequency distribution in Experiment 2 suggest that there was a wide range of the minimum number of moves required to complete the puzzle based on the starting configuration. However, despite the wide range of minimum number of moves based on starting configuration, data patterns for the duration of puzzle solving remained consistent between participants despite the variable starting configurations.

The modifications resulted in a negative predictive value of .93 on the Traditional PDC-HS, indicating that 93% of barriers identified are indeed true barriers. Furthermore, the correlation between the percentage of “No” responses on the first and second administrations of the Traditional PDC-HS more closely resembled the expected correlation discussed in Experiment 1 (see Figure 19). However, discrepant results were obtained for participants in the Digital PDC-HS group, requiring a closer examination of all data collected to inform understanding of the stimulus control of participants’ self-reported responses on the PDC-HS.

In an effort to examine the differences in the percentage of “No” responses between the two versions of the PDC-HS, the investigator analyzed the verbal scores using the stimulus control of the background information and checklist, participant performance, and experienced contingencies for participants in the Traditional PDC-HS group using the same procedures from Experiment 1.

The domain manipulation conditions may have affected participant responding differently. However, differences in responding could also be due to the effect of the verbal behavior of question delivery. Eighty-six percent of the rephrased questions were the same between both participants in the Traditional PDC-HS group (i.e., question 3 of *Task Clarification and Prompting* was rephrased for both participants), however, the rephrasing generated different results. Therefore, analyses need to continue with a larger sample size to fully understand the impact of the verbal behavior of the investigator. However, due to the uncertainty of its effect on participant responding, this should be controlled for in future studies aiming to validate the Digital PDC-HS.

Duration of PDC-HS administration was examined next to gauge whether there were any between-participant differences in speed of responding to PDC-HS questions. However, the mean duration of PDC-HS administration (range 238 s to 292 s) was similar across all participants during both administrations.

Reactivity by participants (Kazdin, 1979) could also be responsible for the differences in participant responding between the Digital and Traditional PDC-HS in Experiment 2. Reactivity is more common when the person being observed is not only aware of the observer's presence, but also their purpose (Kazdin, 1979). Reactivity to the investigator's presence could have created an establishing operation for "correct" answers. To experimentally examine reactivity effects, a Digital PDC-HS could be administered to two groups: 1) one with the observer visible and 2) one with the observer out of the room. By administering the same assessment and changing the visibility of the investigator, differences in participant responding due to investigator presence can be assessed.

Another possibility for the inconsistent PDC-HS results, and the most obvious limitation of the study, is the small sample size. The limitation exists due to time constraints and attrition. Forty-seven individuals showed preliminary interest. Of the 47 who showed interest, 31 participants scheduled a session. Of the 31 participants who scheduled, only 25 showed up to their scheduled sessions. While there is precedent having only two participants in a study employing a multiple baseline design (Hess, 2019; Smith & Wilder, 2018; Wilder et al., 2018), results should still be interpreted with caution. Gast et al. (2014) suggested that to demonstrate true experimental control, the study should have at least three or four participants to determine a functional relation. The procedure should continue to be modified before this experiment is replicated with a larger sample size. The value of the second experiment is in its demonstration of further development of a procedure to assess the technical adequacy of a digital version of the PDC-HS.

General Discussion

The present study attempted to answer the question: Will a digital version of the PDC-HS accurately identify barriers to fluent task completion? There were two aims in the present study: 1) use a complex behavior to validate the Digital PDC-HS; and 2) examine the extent to which the domain manipulations controlled participant responding on the PDC-HS.

In two experiments, 16 participants were tasked with solving a 3x3 sliding puzzle and completing the PDC-HS to identify barriers to fluent puzzle completion. The barriers removed and imposed in the domain manipulation condition in Experiments 1 and 2 were designed to evoke "No" responses from participants to questions in only the *Training* domain, indicating insufficient training was responsible for the lack of fluent puzzle completion. In the barrier removal condition, the repeated success of BST at improving fluent puzzle solving suggested that lack of training was indeed the barrier.

In Experiment 1, while participants did identify *Training* as the primary barrier to fluent puzzle completion in their responses to questions on the second administration of the PDC-HS, other domains were endorsed as well. In Experiment 1, participants identified their written instructions as training, mislabeled materials, did not indicate that other coworkers had received training, and did not acknowledge that there were no competing tasks. Therefore, in Experiment 2, an animated video was presented instead of the written background information and a checklist to help track participant progress was presented instead of the written instructions. Adjusting these two domain manipulations resulted in a difference of responding on the Traditional and Digital PDC-HS relative to Experiment 1.

The present study contributed to the body of literature on the PDC-HS by administering the PDC-HS in a digital format, asking participants who represented employees to respond to the

PDC-HS, and maintaining anonymity of the digital PDC-HS responses. According to the literature, the PDC-HS may be validated in a variety of ways, including: 1) completing a systematic replication of Carr et al. (2013) with an indicated and non-indicated intervention; 2) replicating Wilder et al. (2019) by administering the Digital PDC-HS twice a few weeks apart and determining if results were the same as well as asking participants to identify a predetermined barrier; or, as attempted in the present study, 3) creating an environment with an imposed barrier and comparing the results of the Digital PDC-HS to the results of a "gold standard," the Traditional PDC-HS (previously validated by Bowe & Sellers, 2018; Carr et al., 2013; Ditzian et al., 2015; Wilder et al., 2018).

The second contribution to the literature was administering the PDC-HS to participants who represented employees. While the PDC had been administered to employees (Gravina et al., 2008; Hybza et al., 2013; Miller et al., 2013; Pampino et al., 2004), with the exception of one study by Merritt et al. (2019), past research on the Traditional PDC-HS only utilized supervisor input (Bowe & Sellers, 2018; Carr et al., 2013; Ditzian et al., 2015; Hess, 2019; Smith & Wilder; 2018; Wilder et al., 2018). Merritt and colleagues corroborated employee responses through supervisor input and direct observation, whereas the present study corroborated employee responses using an SD analysis.

A third contribution of the present study is maintaining anonymity of participant responses on the Digital PDC-HS. Previous research has shown that when a third party administers the PDC-HS and the supervisors select and implement interventions based on their interpretation of results, effective behavior change can occur (Bowe & Sellers, 2018; Smith & Wilder, 2018). Merritt et al. (2019) created a precedent for asking employees to respond to the Traditional PDC-HS, which provided insight to other contingencies possibly maintaining tardiness (i.e., dropping a child off at daycare). However, unless a third party interviews employees, employees may not feel comfortable providing feedback to supervisors. Maintaining anonymity of PDC-HS responses is a strength because it may allow for more honest responding (Sellers et al., 2019), however, response anonymity also limits the use of the Digital PDC-HS because it is not possible to ascertain stimulus control of self-reported answers to PDC-HS questions.

Four additional limitations of the present study warrant discussion, an example of which is the undifferentiated responding across domains on the PDC-HS. While the results of prior studies suggest that somewhat undifferentiated responding across domains on the PDC-HS may be common within applied settings (Bowe & Sellers, 2018; Carr et al., 2013; Merritt et al., 2019; Wilder et al., 2018), within a highly controlled and manufactured environment, more rigorous stimulus control over responding is expected. The undifferentiated responding across domains on the PDC-HS decreased in Experiment 2 relative to Experiment 1, especially for participants in the Traditional PDC-HS group, however, these data must be interpreted with caution given the small sample size. Experiment 2 should be replicated with at least four participants in the Traditional PDC-HS group and at least four participants in the Digital PDC-HS group to continue to develop the procedure (Gast et al., 2014). Once the procedure is fully developed, it should be replicated with a large enough sample size where group comparisons are appropriate, as the current sample size is not large enough.

Verbal behavior appears to be a limitation that should be controlled for in future studies. In the present experiment, the verbal behavior of question delivery was not controlled. Therefore, changes in participant responding cannot be solely contributed to the domain manipulations.

Therefore, future experiments should address this by reading the Traditional PDC-HS questions verbatim.

Another limitation is the simulated work environment. Not all domain manipulations exerted the same degree of stimulus control over participant responding (i.e., checklist was not identified as a job aid, participants did not identify that other employees had been trained, and some participants did not identify that feedback was given). While the domain manipulations in Experiment 2 affected participant responding more relative to Experiment 1, they are also a limitation of the present study, as the synthesized and manufactured environment could limit generalization of the results to an applied setting (Poling et al., 2000). To address this limitation, the procedure to validate the Digital PDC-HS should be fully developed, and then the Digital PDC-HS can be administered in a naturalistic setting.

Finally, a variable that could have affected participant responding was that all participants were recruited from undergraduate courses in behavior analysis. Therefore, the study could have cued recall of behavior analytic terms, which could have aided participant understanding of the questions. Tulving and Pearling (1966) define cued recall as having a presentation of a retrieval cue aiding in recall of otherwise inaccessible information. In lay terms, inaccessible information is defined as not remembering something. Inaccessible information becomes accessible when stimuli or cues occur at the time of recall. Studying the possibility of cue dependency could be done by administering the PDC-HS to individuals who had not taken any behavior analysis coursework and completing an SD analysis of their responses to questions on the PDC-HS. This approach would be beneficial because it would help determine if domain manipulations or cue dependency controlled participant responding.

Another possible area of research involves continuing to refine the Digital PDC-HS. While self-report can be useful, it is critical to confirm self-reported responses through direct observation or a secondary observer (Critchfield et al., 1998). Carr et al. (2013) indicated with an asterisk seven questions on the PDC-HS that need to be answered via direct observation. Therefore, to maintain the integrity of the Traditional PDC-HS and maintain anonymity, direct observations should be conducted on a group level and the only questions presented to a participant or an employee should be the 13 questions that do not require direct observation, allowing responses of individual respondents to remain anonymous.

In the present study, to determine the stimulus control maintaining participant self-reported responses on the PDC-HS and to quantify the technical adequacy of the PDC-HS, an SD analysis was completed. As the PDC-HS identifies barriers to task completion by the percentage of questions answered "No" (Carr et al., 2013), specificity and negative predictive value may be of greatest interest (Treat & Viken, 2012). As specificity increases, the accuracy of the PDC-HS identifying a necessary component to task completion missing, or a barrier, increases. As NPV increases, the likelihood that when a barrier is identified on the PDC-HS it is actually a true barrier increases. Identifying false negatives, or false barriers, could result in higher costs due to unnecessary interventions being implemented. While Mager and Pipe (1984) suggested introducing new interventions until effective interventions are found and implemented, ineffective interventions can become expensive and increased costs could inhibit the hiring of behavior analysts in the work setting (Austin, 2000). It is important to maintain the cost effectiveness of the PDC-HS when administering it digitally (Wilder & Hodges, 2018), and SD analyses could aid in assessing the extent to which the Digital PDC-HS is at least as cost effective as the Traditional PDC-HS. However, the Digital PDC-HS may have a greater value than the Traditional PDC-HS given that it is scalable to large organizations and can maintain respondent anonymity. However, as the Digital PDC-HS is used with larger organizations, it should continue to be assessed with an SD analysis to ensure the technical adequacy is maintained as the number of respondents increases.

Finally, the present study contributes to the conversation started by Wilder et al. (2020) regarding how to determine barriers to employee performance. The PDC-HS is designed with 20 questions asking about 20 different topics that can be divided into four domains. Therefore, any question that is answered “No” is considered to be an opportunity for intervention. However, only asking one question per intervention opportunity, may produce misleading results. For example, Carr et al. (2013) stated that *Training* as well as *Performance Consequences, Effort, and Competition* were the barriers to cleaning behaviors. At least one question in each domain was answered “No” indicating that there may have been barriers in multiple domains. Increasing the number of questions per domain could aid in the development of a cut score, or a threshold, for a domain to be identified as a barrier may be another area for future research. For example, a receiver operating characteristic analysis (Pintea & Moldovan, 2009) could be completed to determine an optimal cut score, which could result in a maximization of both sensitivity and specificity for a variety of behavioral assessments (Kerns et al., 2015). Repeating this procedure on the PDC-HS could help determine the optimal sensitivity and specificity of the instrument, which could create an instrument that is the most cost effective.

To create a cut score, multiple questions about the same topic may need to be asked. For example, “Have you received any training? Please explain the training procedures. When were you trained? Who trained you?” However, adding more questions could increase the response effort involved in administering and scoring the PDC-HS, particularly if some questions (such as the examples above) are not answerable with a yes or no response. The present study provides a description of how to quantify the technical adequacy of a modified PDC-HS and describes how to calculate measures such as specificity and NPV, which could be used to develop cut scores and examine the effects of adding additional questions to the assessment.

The first aim of the present study was to use a complex behavior to validate the Digital PDC-HS. Puzzle solving was deemed to be a complex behavior, as different components of solving the puzzle (algorithms) within the learner's repertoire could be applied to solve a novel problem (new configuration of a puzzle). After participants received training, their duration of puzzle solving reliably decreased. Participants also repeatedly identified *Training* as the primary barrier to task completion. Taken together, these results suggest that a complex behavior was used to begin to validate the Digital PDC-HS. Furthermore, to demonstrate validity, reliability had to be demonstrated. Test-retest reliability of the Digital PDC-HS was demonstrated for the Baseline Control Group as responses were similar under similar conditions. However, construct validity has not fully be demonstrated. There are multiple explanations for the different results, and therefore, inferences should be made with extreme caution.

The second aim was to examine the extent to which the domain manipulations controlled participant responding on the PDC-HS. An SD analysis was completed to determine the extent to which the domain manipulations affected participant responding. Based on the frequency of false positive and false negative responses on the PDC-HS in Experiment 1, changes were made to the domain manipulations in Experiment 2. The results of Experiment 2 suggest that the modified domain manipulations controlled participant responding more consistently than the domain manipulations they replaced in Experiment 1, however, further modifications are required to ensure that participant responding is in accordance with the domain manipulations. As the procedure is still being developed, only first steps were taken towards comparing the Traditional and Digital PDC-HS using the "gold standard" method. Once the procedure is fully developed, as confirmed by accurate identification of barriers by responses on the Traditional PDC-HS, a formal between-groups comparison of the Traditional and Digital PDC-HS can be conducted. Although tentative, the preliminary analyses reported here in suggest using the "gold standard" method with SD analyses to corroborate the stimulus control of participant responding may be beneficial for validating the Digital PDC-HS.

The present study continues the evaluation of the PDC-HS, but more importantly, documents the development of a procedure to validate the Digital PDC-HS. If effective, this procedure could be applied to other diagnostic assessments such as the PDC-Parent or the PDC-Safety. After the Digital PDC-HS has been fully developed, it could be used to obtain feedback from a large number of employees, hopefully yielding interventions that are maximally effective within human service settings.

References

Abernathy, B. (2014). Beyond the skinner box: The design and management of organization-wide performance systems. *Journal of Organizational Behavior Management, 34*(4), 235-254. https://doi.org/ 10.1080/01608061.2014.973631

Amigo, S., Smith, A., Ludwig,T. (2008) Using task clarification, goal setting, and feedback to degrease table busing times in a franchise pizza restaurant. *Journal of Organizational Buisness Management, 28*(3), 176-187. doi: 10.1080/0160806082251106

Austin, J. (2000). Performance analysis and performance diagnostics. In J. Austin & J. E., Carr (Eds.), *Handbook of applied behavior analysis* (pp. 321-349). Context Press.

Austin, J., Weatherly, N. L., Gravina, N. E. (2005) Using task calrification, graphic feedback, and verbal feedback to increase closing-task completion in a privately owned restaurant. *Journal of Applied Behaivor Analysis, 38*(1), 117-120. doi: 10.1901/jaba.2005.159-03

Ayllon, T., & Azrin, N. H. (1964). Reinforcement and instructions with mental patients. *Journal of the Experimental Analysis of Behavior, 7*(4), 327-331. https://doi.org/ 10.1901/jeab.1964.7-327

Beal, S., & Eubanks, J. (2003). Self-report bias and accuracy in a simulated work setting: effects of combined feedback on task and self-reported performance. *Journal of Organizational Behavior Management*, *22*(1), 3-31. https://doi.org/ 10.1300/J075v22n01_02

Bijou, S. W., Peterson, R. F., & Ault, M. H. (1968). A method to integrate descriptive and experimental field studies at the level of data and empirical concepts. *Journal of Applied Behavior Analysis, 1*(2), 175–191. https://doi.org/10.1901/jaba.1968.1-175

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel, A. M. (2003). Enhanced pitch sensitivity in individuals with autism: A signal detection analysis. *Journal of Cognitive Neuroscience*, *15*(2), 226-235. https://doi.org/ 10.1162/089892903321208169.

Bowe, M., & Sellers, T. P. (2018). Evaluating the Performance Diagnostic Checklist-Human Services to assess incorrect error-correction procedures by preschool paraprofessionals. *Journal of Applied Behavior Analysis*, *51*(1), 166-176.https://doi.org/10.1002/jaba.428

Bucklin, B. R., Alvero, A. M., Dickinson, A. M., Austin, J., & Jackson, A. K. (2000). Industrial organizational psychology and organizational behavior management: An objective comparison. *Journal of Organizational Behavior Management, 20*(2), 27-75. https://doi.org/10.1300/j075v20n02_03

Carr, J. E., Wilder, D., Majdalany, L., Mathisen, D., & Strain, L. (2013). An assessment based solution to a human-service employee performance problem: An evaluation of the Performance Diagnostic Checklist-Human Services. *Behavior Analysis in Practice, 6*(1),16-32. https://doi.org/10.1007/BF03391789

Carr, J. E., & Wilder, D. (2016). The Performance Diagnostic Checklist - Human Services: A correction. *Behavior Analysis in Practice, 9*(1), 63. https://doi.org/10.1007/s40617-015-0099-3

Critchfield, T. (1993). Signal-detection properties of verbal self-reports. *Journal of the Experimental Analysis of Behavior, 60*(3), 495-514. https://doi.org/10.1901/jeab.1993.60-495

Critchfield T. S., Tucker, J. A., & Vuchinich, R. E. (1998). Self-report methods (pp. 471-508) in K. A. Lattal & M. Perone (Eds.) Handbook of research methods in human operant behavior. Plenum Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests.*Psychological Bulletin*, *52*(4), 281-302. https://doi.org/10.1037/h0040957

Daniels, A. C., & Bailey, J. S. (2014). *Performance management: Changing behavior that drives organizational performance*. Georgia: Performance Management Publications.

Ditzian, K., Wilder, D. A., King, A., & Tanz, J. (2015). An evaluation of the performance diagnostic checklist - human services to assess an employee performance problem in a center-based autism treatment facility. *Journal of Applied Behavior Analysis*, *48*(1), 199-203. https://doi.org/10.1002/jaba.171

Doll, J., Livesey, J., McHaffie, E., & Ludwig, T. D. (2007). Keeping an uphill edge: Managing cleaning behaviors at a ski shop. *Journal of Organizational Behavior Managemen*t, *27*(3), 41-60. doi:10.1300/J075v27n03_04

Dye, C., Watt, C. J., Bleed, D. M., Hosseini, S. M., & Raviglione, M. C. (2005). Evolution of tuberculosis control and prospects for reducing tuberculosis incidence, prevalence, and deaths globally. *Journal of the American Medical Association*, *293*(22), 2767–2775. https://doi.org/10.1001/jama.293.22.2767

Eikenhout, N., & Autsin, J. (2005). Using goals, feedback, reinforcement, and a performance matrix to improve customer service in a large department store. *Journal of Organizational Business Management, 24*(3), 27-62. doi: 10.1300/J075v24n03_02

Fante, R., Gravina, N., Betz, A., & Austin, J. (2010). Structural and treatment analyses of safe and at- risk postures performed by pharmacy employees. *Journal of Organizational Behavior Management*, *30*(4), 325-338. https://doi.org/10.1080/01608061.2010.520143

Fisher, W.W., Pawich, T.L., Dickes, N., Paden, A.R., & Toussaint, K. (2014). Increasing the saliency of behavior-consequence relations for children with autism who exhibit persistent errors. *Journal of Applied Behavior Analysis*, *47*(4), 738-748. https://doi.org/10.1002/jaba.172

Floyd, R. G., Phaneuf, R. L., & Wilczynski, S. M. (2005). Measurement Properties of Indirect Assessment Methods for Functional Behavioral Assessment: A Review of Research. *SchoolPsychology Review*, *34*(1), 58–73. Retrieved from http://search.ebscohost.com.www2.lib.ku.edu/login.aspx?direct=true&db=aph&AN=16608418&site=ehost-live

Gast, D.L., Lloyd, B.P, & Ledford, J.R. (2014) Multiple baseline and multiple probe designs. In D.L. Gast, B.P. Lloyd, & J.R. Ledford (Eds.) *Single case research methodology: Applications in special education and behavioral sciences*. New York.

Gravina, N., VanWagner, M., & Austin, J. (2008). Increasing physical therapy equipment preparation using task clarification, feedback, and environmental manipulations. *Journal of Organizational Behavior Management, 28*(2), 110-122. https://doi.org/10.1080/01608060802100931

Gresham, F. M. (2003). Establishing the technical adequacy of functional behavioral assessment: conceptual and measurement challenges. *Behavioral Disorders, 28*(3), 282–298. https://doi.org/10.1177/019874290302800305

Hess, B. H. (2019). Evaluating the Performance Diagnostic Checklist-Human Services to treat performance problems of adults with intellectual disabilities. (Unpublished master's thesis). Utah State University, Logan, Utah.

How Kids Make Money. (2018, January 27). *Solving a 3x3 sliding puzzle.* [Video file]. Retrieved from https://www.youtube.com/watch?v=SJjzDyNCAFQ

Hybza, M. M., Stokes, T. F., Hayman, M., & Schatzberg T. (2013). Increasing medicaid revenue generations for services by school psychologists. *Journal of Organizational Behavior Management*, *33*(1), 55-67. https://doi.org/10.1080/01608061.2012.758011

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1982, 1994). Toward a functional analysis of self injury. *Journal of Applied Behavior Analysis, 27*, 197-209. (Reprinted from *Analysis and Intervention in Developmental Disabilities, 2*, 3-20, 1982)

Iwata, B. A., DeLeon, I. G., & Roscoe, E. M. (2013). Reliability and validity of the functional analysis screening tool. *Journal of Applied Behavior Analysis, 46*(1), 271-284. https://doi.org/10.1002/jaba.31

Kazdin, A. E. (1979). Unobtrusive measures in behavioral assessment. *Journal of Applied Behavior Analysis*, *12*(4), 713–724. https://doi.org/10.1901/jaba.1979.12-713

Kazdin, A. E. (1999). Drawing valid inferences I: Internal and external validity. In A. E. Kazdin (Ed.) *Research design in clinical psychology* (3rd ed.) (pp. 15-39). Allyn and Bacon.

Kerns, C. M., Maddox, B. B., Kendall, P. C., Rump, K., Berry, L., Schultz, R., Souders, M. C., Bennett, A., Herrington, J. Miller, J. (2015). Brief measures of anxiety in non-treatment-seeking youth with autism spectrum disorder. *Autism, 19*(8), 969-979. https://doi.org/ 10.1177/1362361314558465

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M. & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Lebbon, A., Austin, J., Rost, K., & Stanley, L. (2011). Improving safe consumer transfers in a day treatment setting using training and feedback. *Behavior Analysis in Practice, 4*(2), 35-42. doi: 10.1007/BF03391782

Lerman, D. C., Tetreault, A., Hovanetz, A., Bellaci, E., Miller, J., Karp, H., Mahmood, A., Strobel, M., Mullen, S., Keyl, A., & Toupard, A. (2010). Applying signal-detection theory to the study of observer accuracy and bias in behavioral assessment. *Journal of Applied Behavior Analysis*, *43*(2), 195–213. https://doi.org/10.1901/jaba.2010.43-195

Lippman, L. G. & Meyer, M. E. (1967). Fixed-interval performance as related to instructions and to subjects' verbalizations of the contingency. *Psychonomic Science, 8*(4), 135-136. https://doi.org/ 10.3758/BF03331586

Loughrey, T. O., Marshall, G. K., Bellizzi, A., & Wilder, D. A. (2013). The use of video modeling, prompting, and feedback to increase credit card promotion in a retail setting. *Journal of Organziational Buisness Management*, *33*(3), 200-208. doi: 10.1080/01608061.2013.815097

Martinez-Onstott, B., Wilder, D., & Sigurdsson, S. (2016). Identifying the variables contributing to at-risk performance: Initial evaluation of the Performance Diagnostic Checklist – Safety (PDC-Safety). *Journal of Organizational Behavior Management*, *36*(1), 80-93. https://doi.org/10.1080/01608061.2016.1152209

McIntosh, K., Borgmeier, C., Anderson, C. M., Rodrigues, B. J., Tobin, T. J. (2008). Technical adequacy of the functional assessment checklist: Teachers and staff (FACTS) FBA interview measure. *Journal of Positive Behavior Interventions, 10*(1), 33-45. https://doi.org/10.1177/10983007311619

McSween, T. (2003). The values-based safety process: Improving your safety culture with behavior-based safety (2nd ed.). Hoboken, NJ: Wiley. Retrieved from https://epdf.pub/the-values-based-safety-process-improving-your-safety-culture-with-behavior-base.html

Merritt, T. A. (2017). An Evaluation of the Performance Diagnostic Checklist-Human Services on the Timeliness of Employees in a School for Students with Autism Spectrum Disorder (Unpublished doctoral dissertation). University of Kansas, Lawrence, KS

Merritt, T. A., DiGennaro Reed, F. D., & Martinez, C. E. (2019). Using the Performance Diagnostic Checklist–Human Services to identify an indicated intervention to decrease employee tardiness. *Journal of Applied Behavior Analysis, 52*(4), 1034-1048. https://doi.org/10.1002/jaba.643

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749. http:// doi.org /10.1037/0003-066X.50.9.7

Miller, M. V., Carlson, J., Sigurdsson, S. O. (2014). Improving treatment integrity in a human service setting using lottery-based incentives. *Journal of Organizational Behavior Management, 34*(1), 29-38. https://doi.org/10.1080/01608061.2013.873381

Mudford, O., Zeleny, J., Fisher, W., Klum, M., & Owen, T. (2011). Calibration of observational measurement of rate of responding. *Journal of Applied Behavior Analysis, 44*(3), 571-586. https://doi.org/10.1901/jaba.2011.44-571

Pampino, R. N., Jr., Heering, P. W., Wilder, D. A., Barton, C. G., & Burson, L. A. (2004). The use of the Performance Diagnostic Checklist to guide intervention selection in an independently owned coffee shop. *Journal of Organizational Behavior Management, 23*(2-3), 5-19. https://doi/org/10.1300/J075v23n02_02

Pampino, R. N., Wilder, D. A., & Binder, C. (2005). The use of functional assessment and frequency building procedures to increase product knowledge and data entry skills among foremen in a construction organization. *Journal of Organizational Behavior Management*, *25*(2), 1-36. https://doi.org/10.1300/J075v25n02_01

Pintea, S. & Moldovan, R. (2009). The receiver-operating characteristic (ROC) analysis: fundamentals and applications in clinical psychology. *Journal of Cognitive and Behavioral Psychotherapies, 9*(1), 49–66. Retrieved from https://www.researchgate.net/profile/Ramona_Moldovan/publication/256454600_The_Receiver-Operating_Characteristic_ROC_analysis_Fundamentals_and_applications_in_clinical_psychology/links/0deec522b97e08e9aa000000/The-Receiver-Operating-Characteristic-ROC-analysis-Fundamentals-and-applications-in-clinical-psychology.pdf

Poling, A., Weetjens, B. J., Cox, C., Beyene, N., Bach, H., & Sully, A. (2010). Teaching giant african pouched rats to find landmines: operant conditioning with real consequences. *Behavior analysis in practice*, *3*(2), 19–25. https://doi.org/10.1007/BF03391761

Poling, A., Weetjens, B., Cox, C., Beyene, N., Durgin, A., & Mahoney, A. (2011). Tuberculosis detection by giant African pouched rats. *The Behavior Analyst*, *34*(1), 47–54. https://doi.org/10.1007/BF03392234

Reed, D. D., & Azulay, R. L. (2011). A Microsoft Excel® 2010 based tool for calculating interobserver agreement. *Behavior analysis in practice*, *4*(2), 45–52. https://doi.org/10.1007/BF03391783

Reinefeld, A. (1993). Complete Solution of the Eight-Puzzle and the Benefit of Node Ordering in IDA. *International Joint Conference on Artificial Intelligence*, 248–253.

Sasson, J. Alvero, A., & Austin, J. (2006). Effects of process and human performance improvement strategies. *Journal of Organizational Behavior Management*, *26*(3), 43-78. https://doi.org/10.1300/J075v26n03_02

Schlinger, H., & Blakely, E. (1987). Function-altering effects of contingency-specifying stimuli. *The Behavior Analyst*, *10*(1), 41–45. https://doi.org/10.1007/BF03392405

Sellers, T. P., Valentino, A. L., Landon, T. J., & Aiello, S. (2019). Board certified behavior analysts' supervisory practices of trainees: Survey results and recommendations. *Behavior Analysis in Practice, 12*(3), 536-546. https://doi.org/10.1007/s40617-019-00367-0

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. Authors Cooperative.

Smith, G. D., Lambert, J. V., Moore, Z. (2013) Behavior description effect on accuracy and reliability. *The Journal of General Psychology*, *140*(4), 269-281. https://doi.org/ 10.1080/00221309.2013.818525

Smith, M., & Wilder, D. A. (2018). The use of the Performance Diagnostic Checklist – Human Services to asses and improve the job performance of individuals with intellectual disabilities. *Behavior Analysis in Practice, 11*(2), 148-153. https://doi.org/10.1007/s40617-018-0213-4

Skinner, B. F. (1969). *Contingencies of reinforcement: A theoretical analysis*. N.J: Prentice Hall

Skinner, B. F. (1977). Herrnstein and the evolution of behaviorism. *American Psychologist, 32*(12), 1006-1012.http://doi.org/10.1037/0003-066X.32.12.1006

Treat, T. A., & Viken, R. J. (2012). Measuring test performance with signal detection theory techniques. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbooks in psychology®. APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics* (p. 723–744). American Psychological Association. https://doi.org/10.1037/13619-038

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*(4), 381-391. https://doi.org/10.1016/S0022-5371(66)80048-8

Turner, L., & Solomon, R. (1962). Human traumatic avoidance learning: Theory and experiments on the operant-respondent distinction and failures to learn, *76*(40), 1-32. https://doi.org/ 10.1037/h0093915

Villacorta, J. (2017). *Evaluation of a tool to identify the variables responsible for poor parent treatment implementation: The Performance Diagnostic Checklist - Parent*. Unpublished master's thesis, Florida Institute of Technology: Melbourne, Florida.

Watson, P. J., & Workman, E. A. (1981). The non-concurrent multiple baseline across individuals design: An extension of the traditional multiple baseline design. *Journal of Behavior Therapy and Experimental Psychiatry, 12*(3), 257-259. https://doi.org/ 10.1016/0005-7916(81)90055-0

Wilder, D. A., & Hodges, A. C. (2018). Performance-management assessment. In B. Wine & J. K. Pritchard (Eds.,), *Organizational behavior management: The essentials*. Hedgehog Publishers.

Wilder, D. A., Lipschultz, J., & Gehrman, C. (2018) An evaluation of the performance diagnostic checklist - human services (PDC-HS) across domains. *Behavior Analysis in Practice*, *11*(2), 129-138. https://doi.org/10.1007/s40617-018-0243-y

Wilder, D., Lipschultz, J., Gehrman, C., Ertel, H., & Hodges, A. (2019). A Preliminary Assessment of the Validity and Reliability of the Performance Diagnostic Checklist-Human Services. *Journal of Organizational Behavior Management*, *39*(3-4), 194-212. https://doi.org/10.1080/01608061.2019.1666772

Wilder, D. A., Cymbal, D., & Villacorta, J. (2020). The Performance Diagnostic Checklist-Human Services: A brief review. *Journal of Applied Behavior Analysis*, *53*(2), 1170-1176. https://doi.org/10.1002/jaba.676

Wustenraad, D. (n.d.). Create interactive puzzles easily. Retrieved May 20, 2020, from http://www.puzzel.org/en

Figure 1.

The sliding puzzle used for the experimental sessions in Experiment 1 and Experiment 2. The top picture shows a starting configuration. The bottom picture shows the solved puzzle.

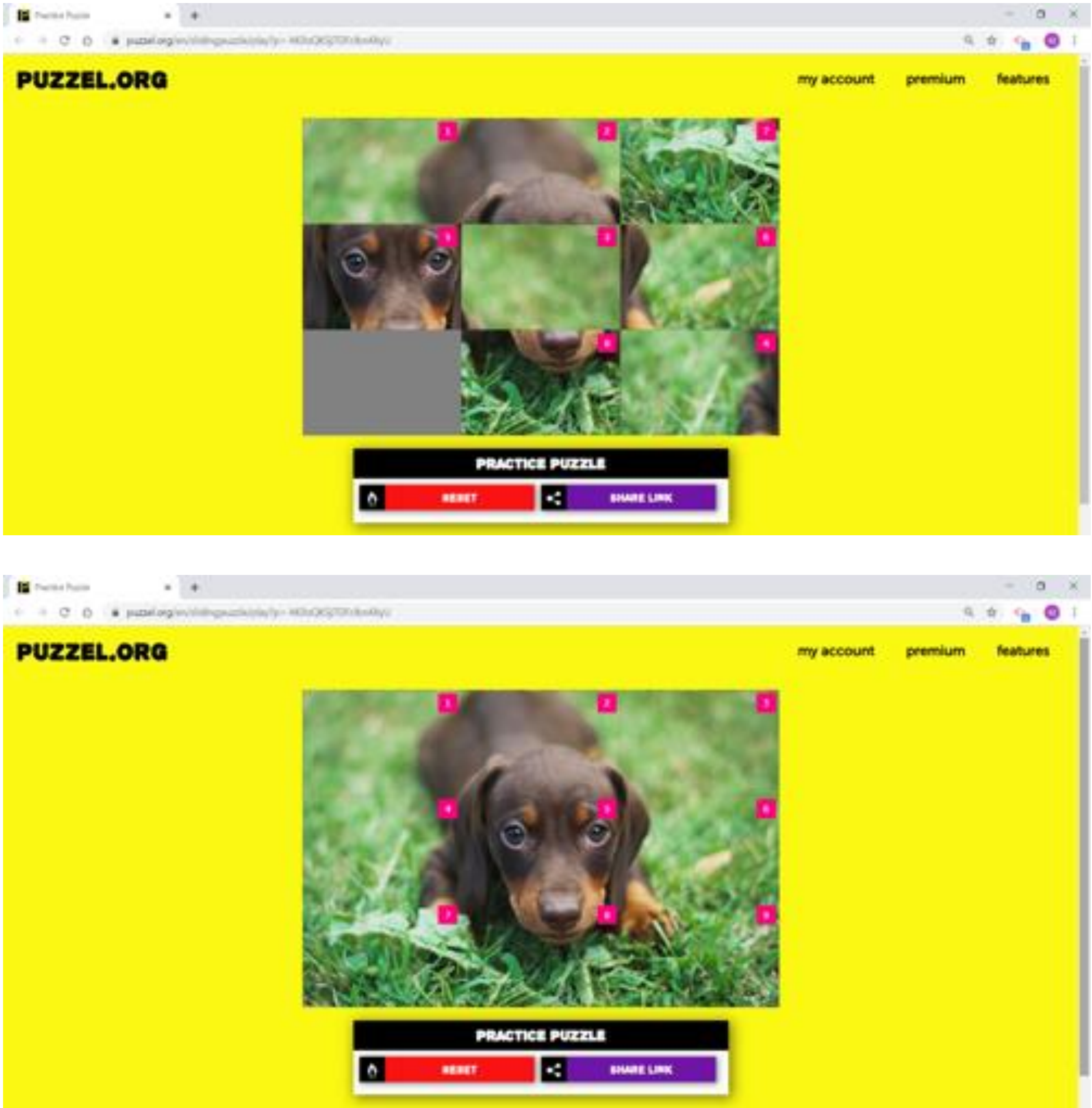

Figure 2.

The sliding puzzle used during BST in both Experiment 1 and Experiment 2. The top picture shows a starting configuration of the puzzle. The bottom picture shows the solved puzzle.

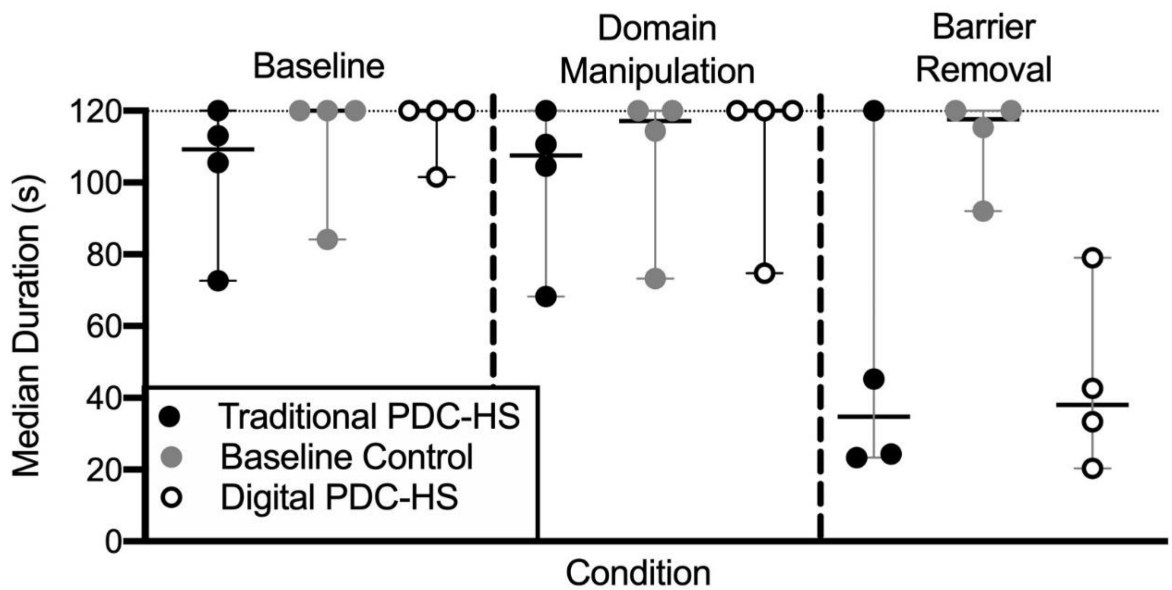

Figure 3

The median duration of puzzle solving for participants in the Traditional PDC-HS (black filled circle), Baseline Control (gray filled circles) and the Digital PDC-HS (white filled circle) groups across all experimental conditions. Each data point represents an individual participant. The dotted line represents 120 s, or the maximum amount of time and the black horizontal line represents the median time. Note. Participants in the Baseline Control Group did not experience domain manipulations or barrier removal. Therefore, Baseline Control represents the duration across the three baseline conditions.

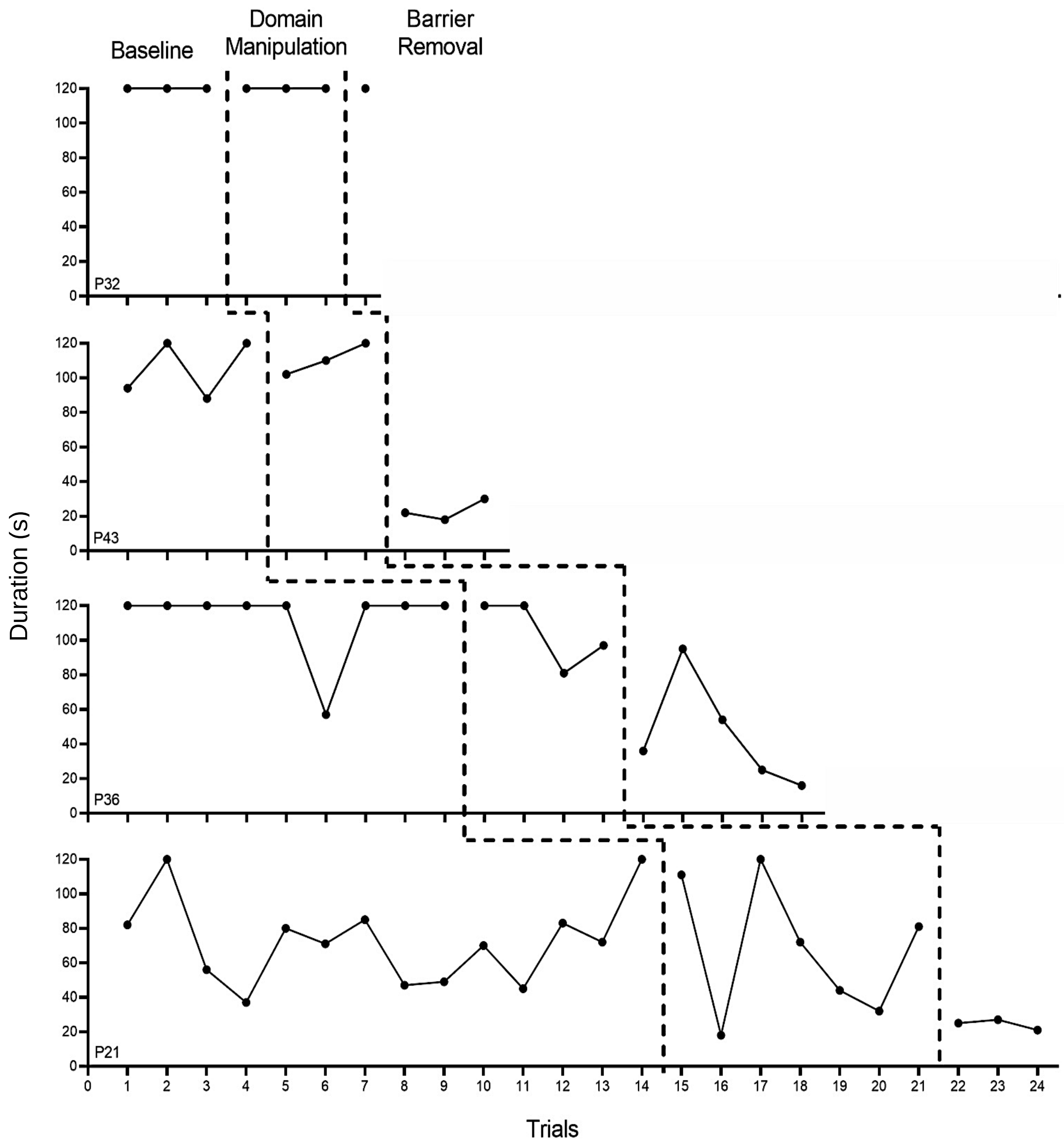

Figure 4.

The duration of puzzle solving for all participants in the Traditional PDC-HS Group in Experiment 1.

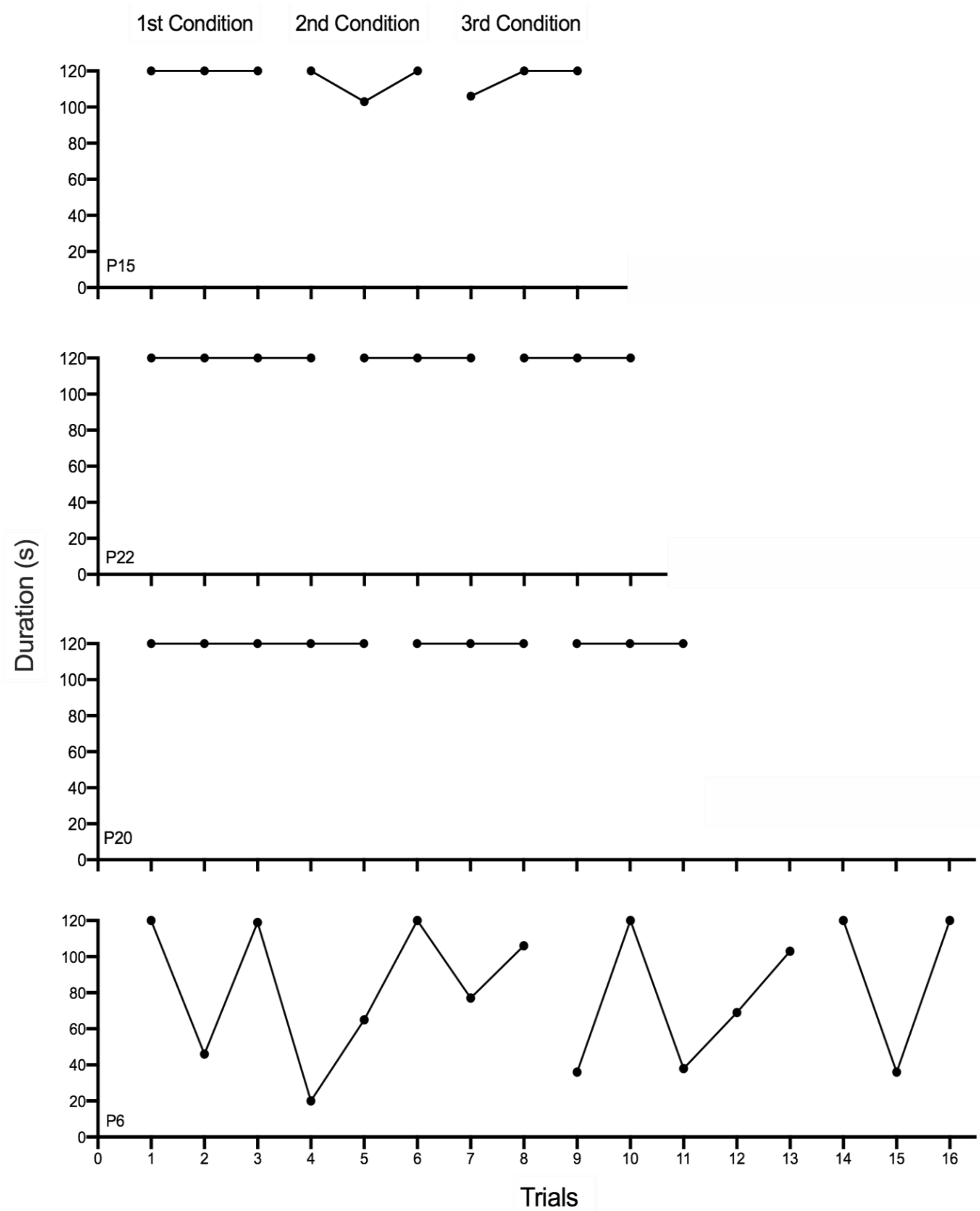

Figure 5.

The duration of puzzle solving for all participants in the Baseline Control Group in Experiment 1.

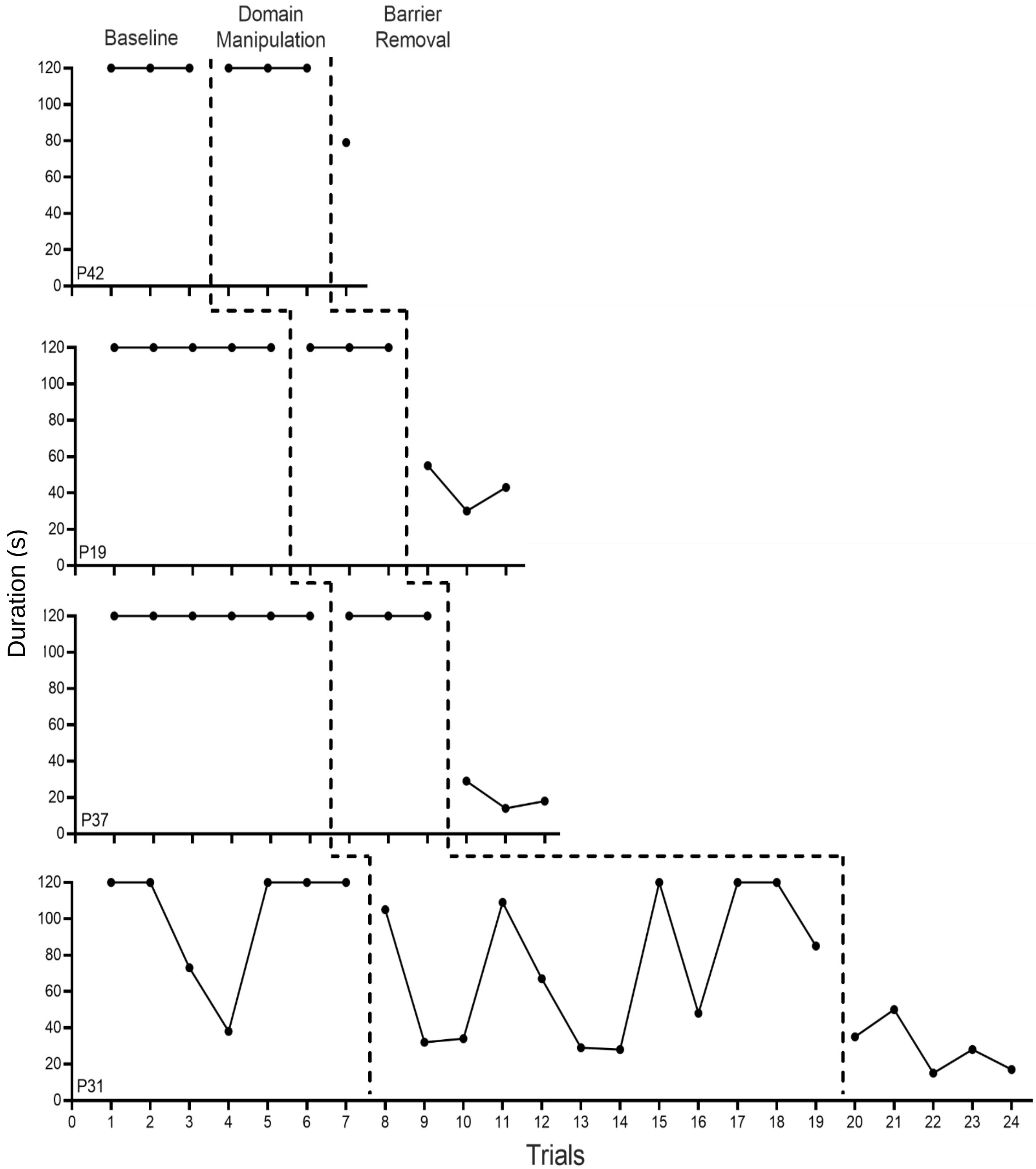

Figure 6.

The duration of puzzle solving for all participants in the Digital PDC-HS Group in Experiment 1.

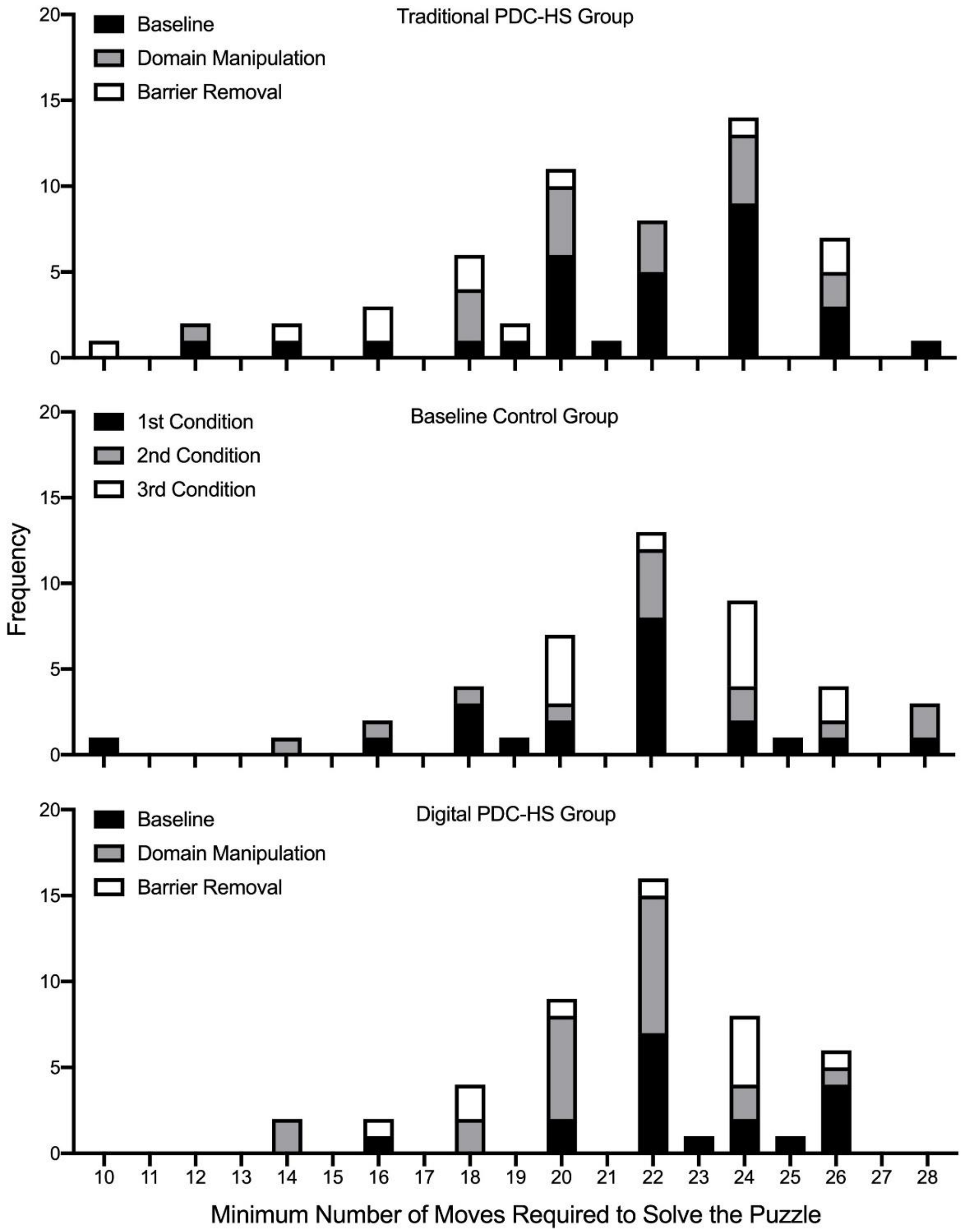

Figure 7.

The frequency distribution of the minimum number of moves required based on the starting configuration in Experiment 1. *Note*. Participants in the Baseline Control Group experienced baseline contingencies across all conditions and did not experience domain manipulations or barrier removal.

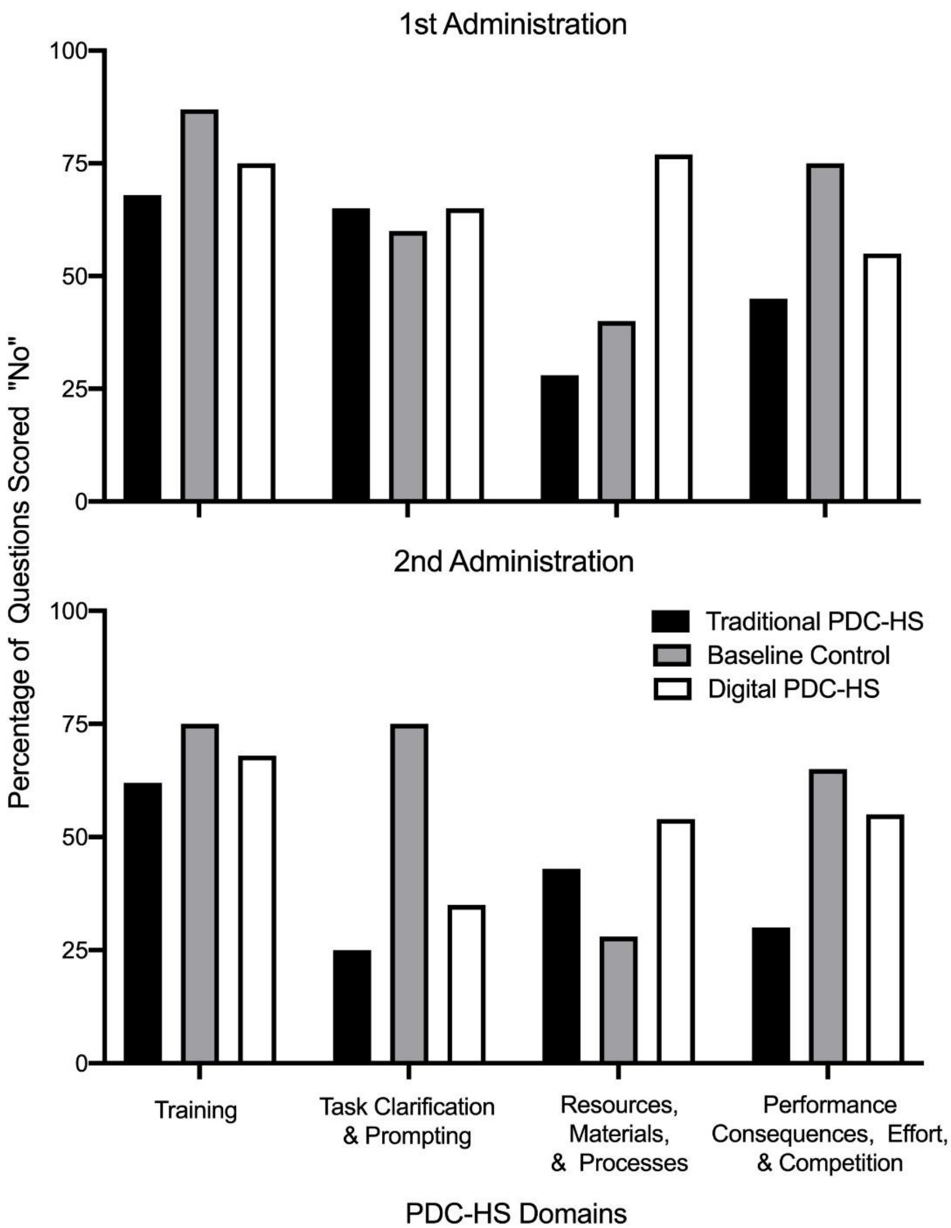

Figure 8.

The results of the PDC-HS across groups in Experiment 1. Black histograms represent the Traditional PDC-HS group, gray histograms represent the Baseline Control group, and the white histograms represent the Digital PDC-HS group. The legend applies to both graphs.

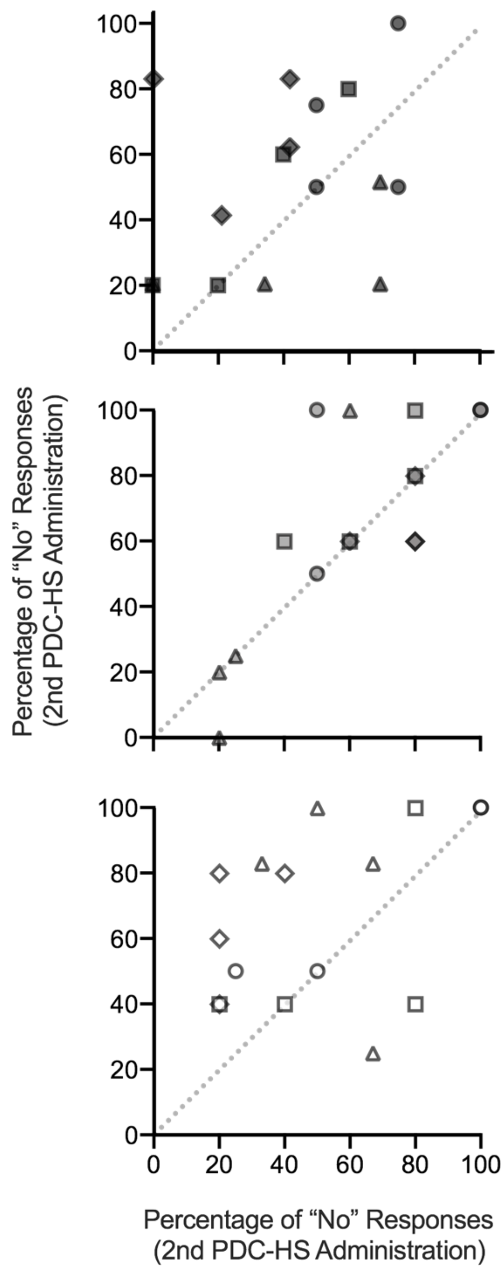

Figure 9.

Correlations between the percentage of "No" responses on the first and second administrations of the PDC-HS across groups in Experiment 1. The Traditional PDC-HS Group is on top, the Baseline Control PDC-HS Group is in the middle, and the Digital PDC-HS Group is on the bottom. The darker the data point indicates more overlapping data points. The legend applies to all graphs.

	True Positive
	True Negative
	False Positive
	False Negative

V	Question was read and answered as the question is written.
R	Question was already paraphrased when the question was given.
U	Follow up information without being asked
C	Information was given when participant manded for it

	Traditional PDC-HS							
	1st Administration				2nd Administration			
	32	43	36	21	32	43	36	21
Training								
1					R	V	V	R
2					V	V	U	V
3					V	V	V	V
4					R	V	V	V
Task Clarification & Prompting								
1					V	V	V	R
2					R	U	U	R
3					R	R	R	R
4					R	R	R	R
5					R	R	R	R
Resources, Materials, & Processes								
1					V	V	V	V
2					V	V	R	V
3					R	R	R	R
4					R	R	R	R
5					V	C	R	R
6					R	R	C	N/A
Performance Consequences, Effort, & Competition								
1					V	V	V	V
2					V	V	R	V
3					V	V	V	V
4					R	R	R	R
5					V	U	C	R

	Digital PDC-HS							
	1st Administration				2nd Administration			
	A	B	C	D	A	B	C	D
Training								
1								
2								
3								
4								
Task Clarification & Prompting								
1								
2								
3								
4								
5								
Resources, Materials, & Processes								
1								
2								
3								
4								
5								
6								
Performance Consequences, Effort, & Competition								
1								
2								
3								
4								
5								

Figure 10.

Tables displaying the SD and Verbal Score Analysis of Experiment 1. The legend applies to both tables. PDC-HS results of Experiment 1 by question while also displaying the SD analysis. Verbal scores were calculated to determine if verbal behavior affected participant responding. Verbal Score definitions for V – Verbatim, R – Rephrase, U – Unprompted, and C – Clarification are in Appendix D.

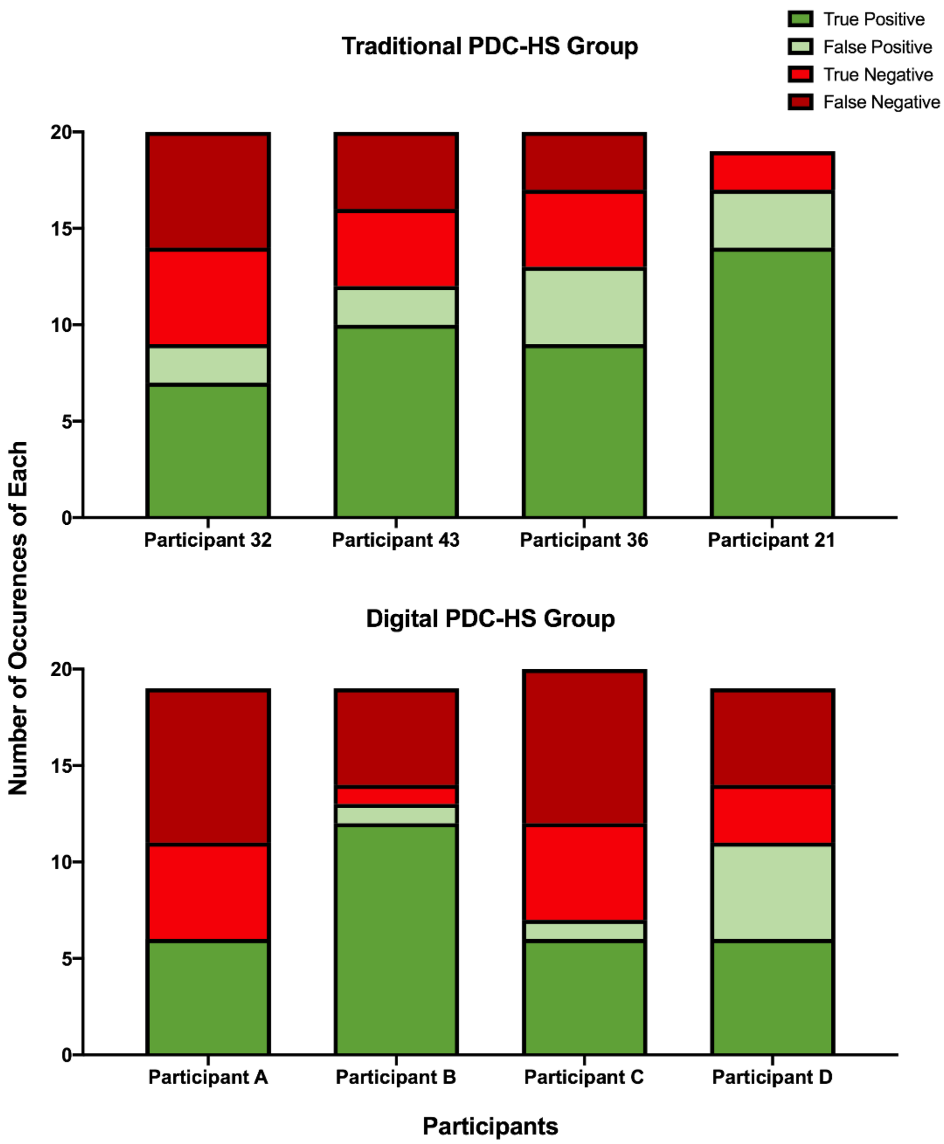

Figure 11.

The number of participant responses in each of the four SD categories (i.e., true positive, false positive, true negative, and false negative). Scaled to the *x*-axis are the participants and scaled the *y*-axis is the number of occurrences of each response. Each histogram represents one participant. Green represents true positives, light green represents false positives, red represents true negatives, and maroon represents false negatives (see Table 2). The legend applies to both graphs.

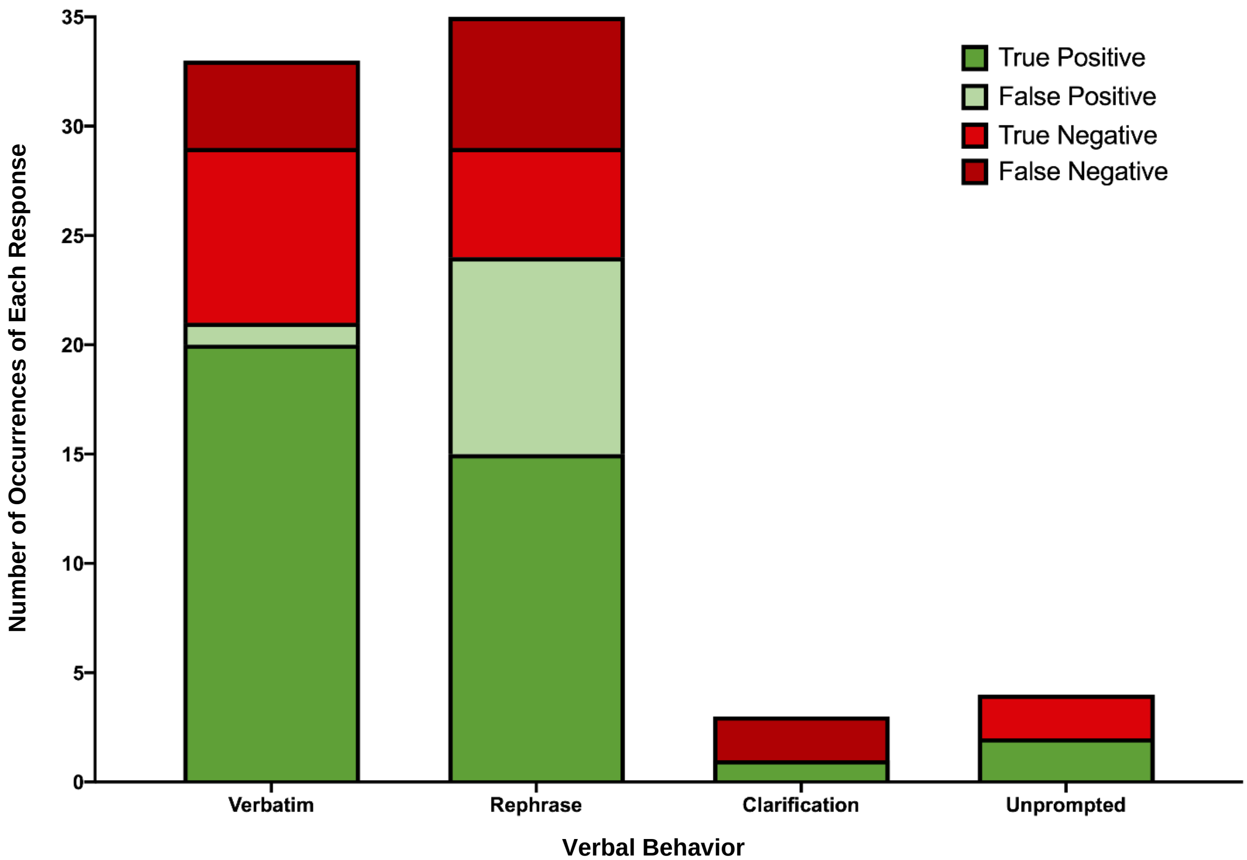

Figure 12.

The verbal behavior of question delivery stratified by the four SD analysis categories for all participants in the Traditional PDC-HS group. Scaled to the *x*-axis is the verbal behavior of question delivery (i.e., verbatim, rephrase, clarification, and unprompted; see Table 2and scaled to the *y*-axis is the number of occurrences of each response. Green represents true positives, light green represents false positives, red represents true negatives, and maroon represents false negatives.

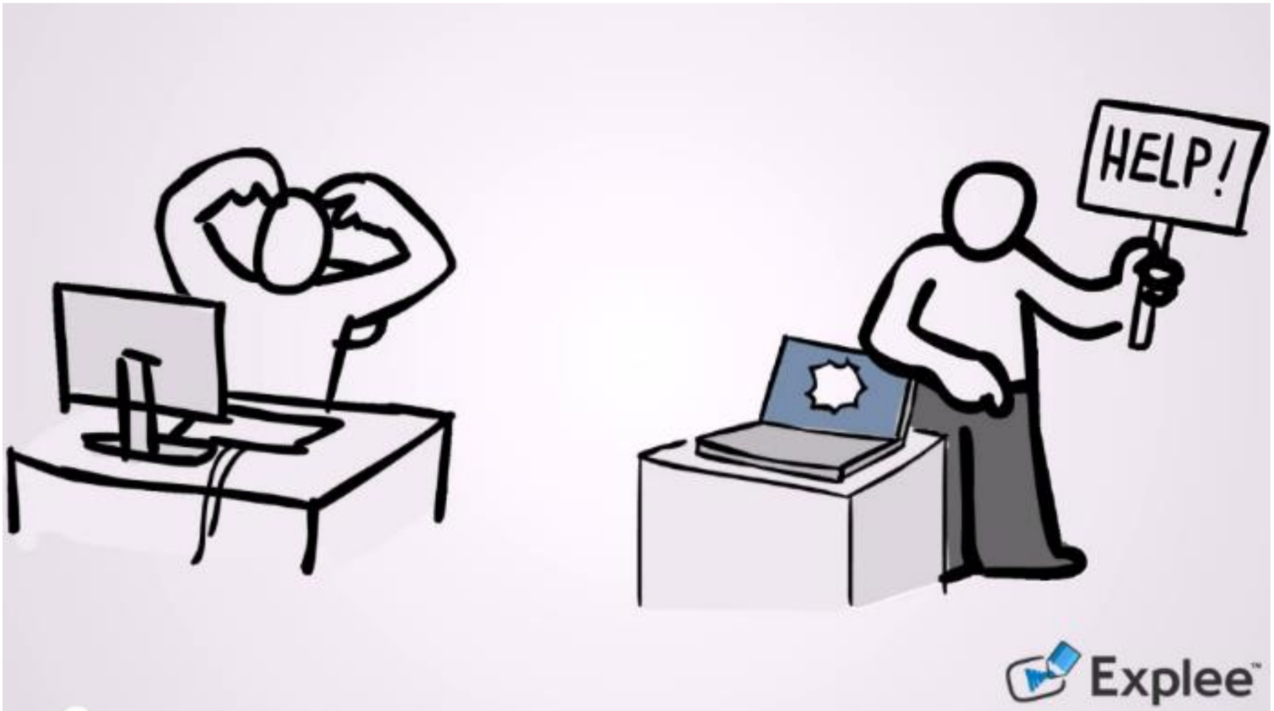

Figure 13.

A screenshot of the animated video that participants were shown in Experiment 2. Video can be viewed at https://explee.com/video/0rpwd95.

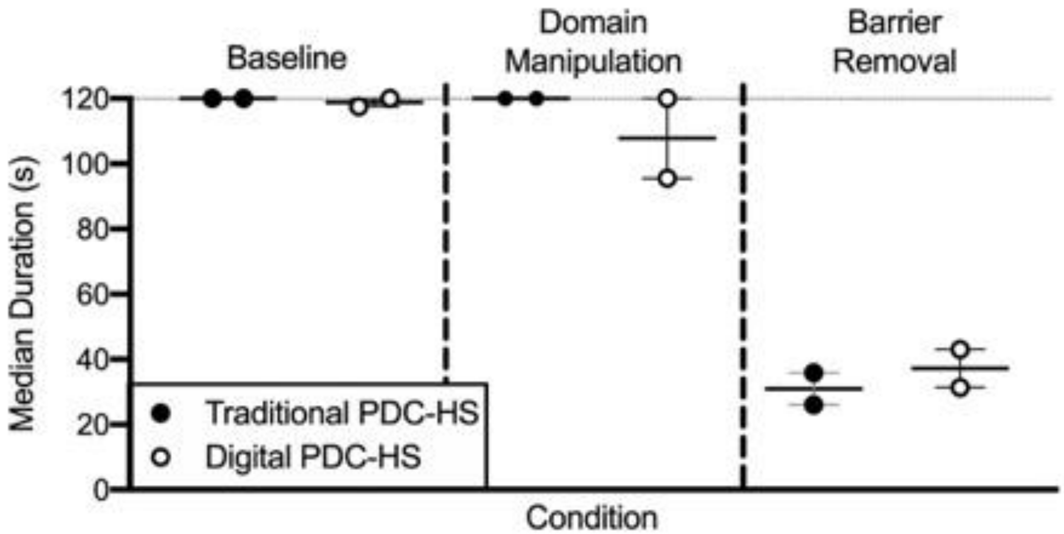

Figure 14.

The median duration of puzzle solving for participants in the Traditional PDC-HS (black filled circle), and the Digital PDC-HS (white filled circle) groups across all experimental conditions. Each data point represents an individual participant. The dotted line represents 120 s, or the maximum amount of time and the black horizontal line represents the median time.

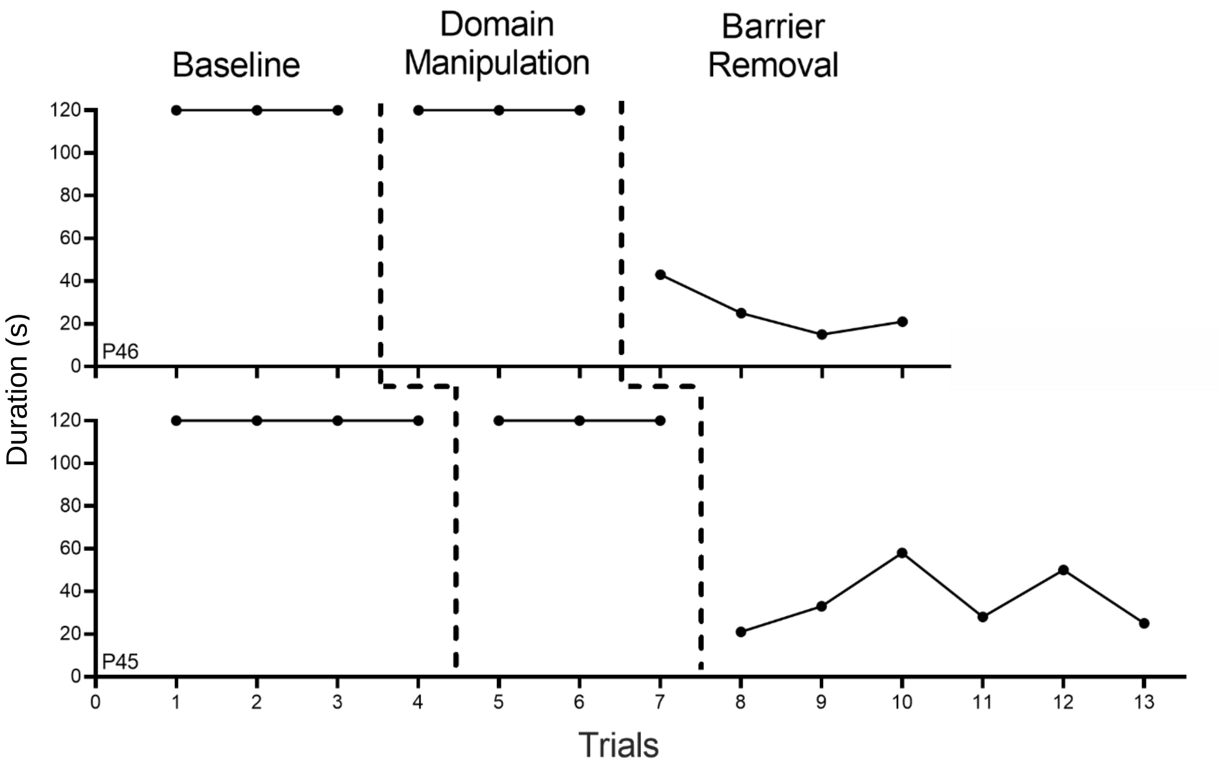

Figure 15.

The duration of puzzle solving for all participants in the Traditional PDC-HS Group in Experiment 2.

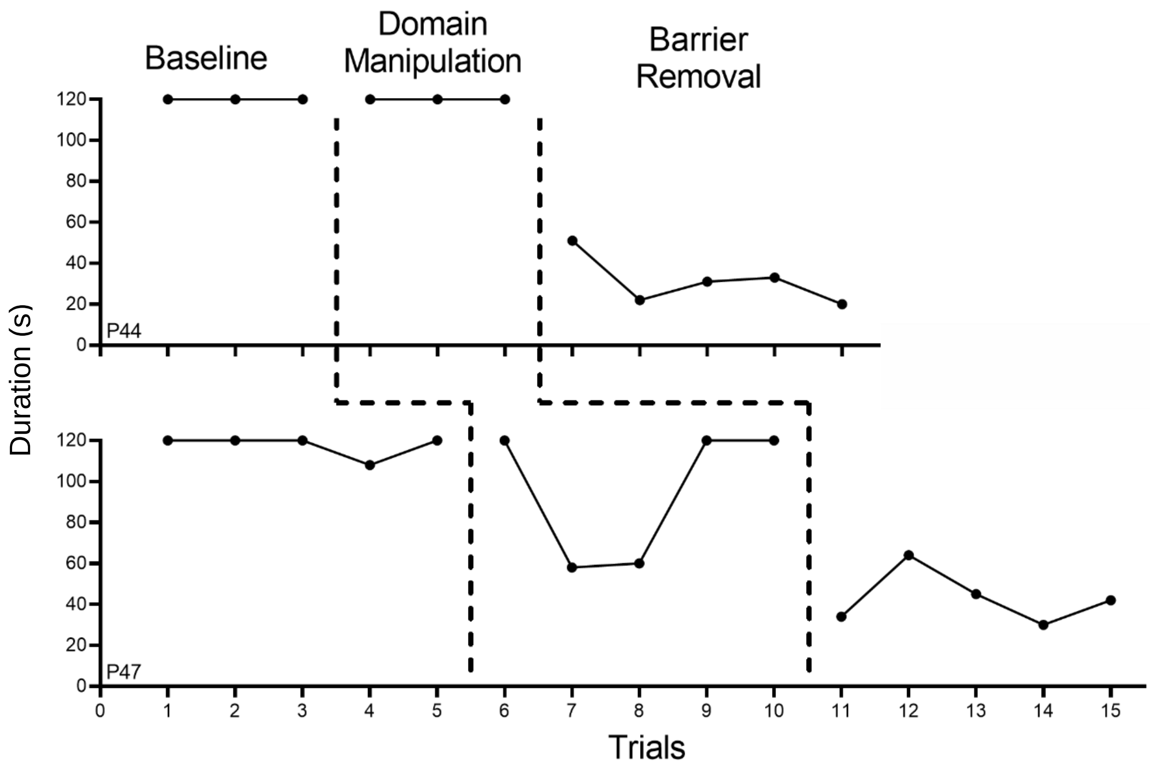

Figure 16.

The duration of puzzle solving for all participants in the Digital PDC-HS Group in Experiment 2.

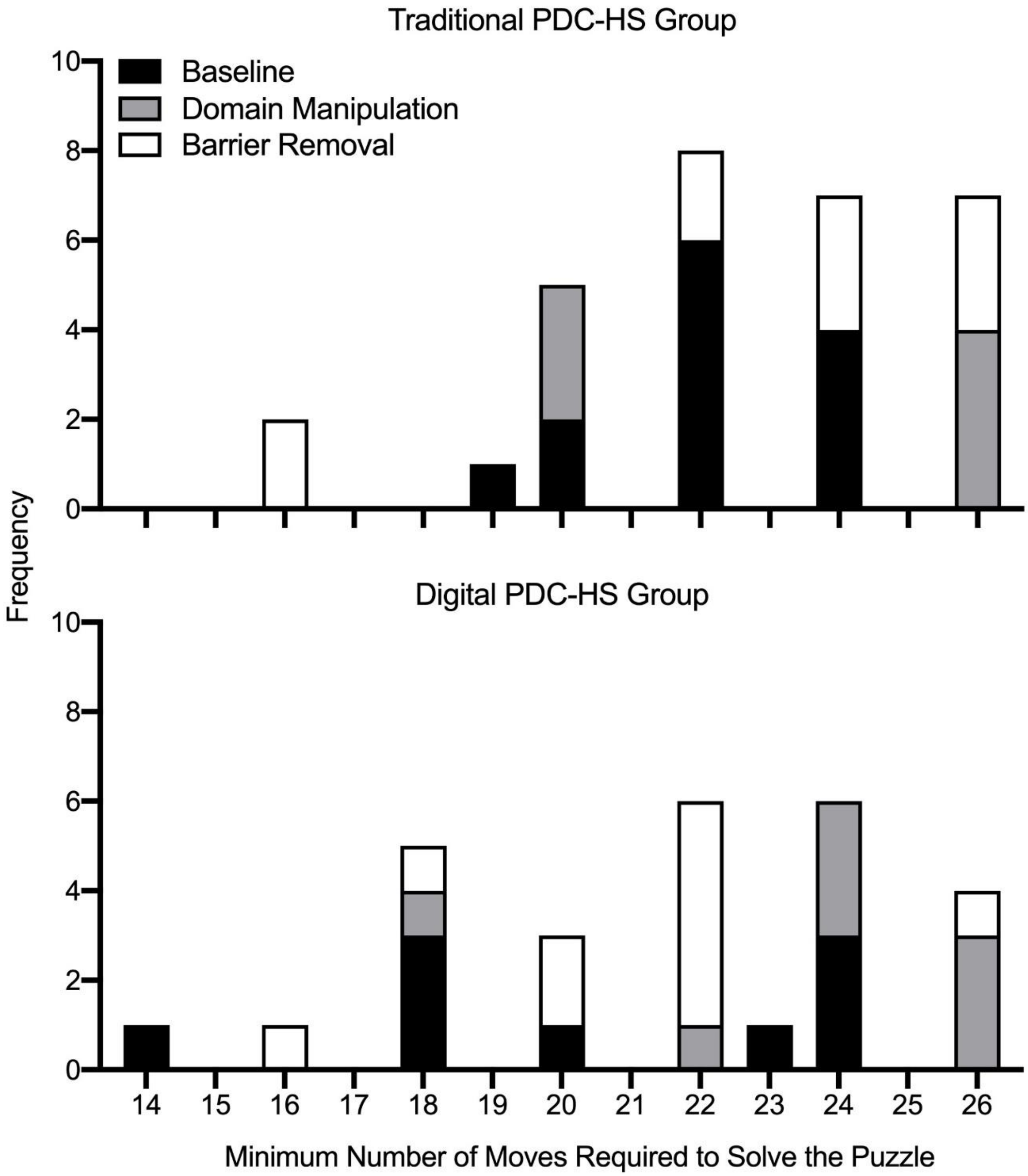

Figure 17.

The frequency distribution of the minimum moves required based on the starting configuration in Experiment 2. The legend applies to both graphs.

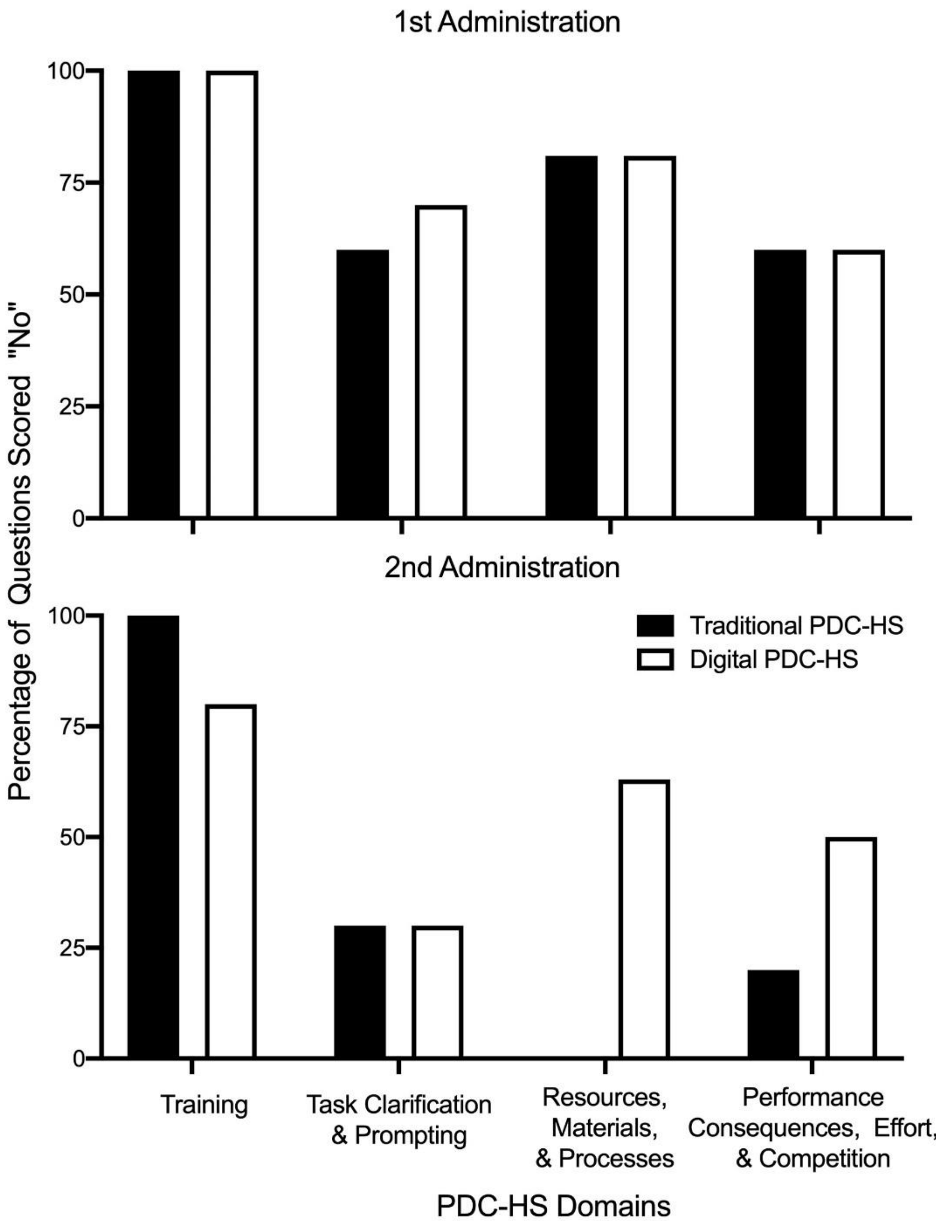

Figure 18.

The results of the PDC-HS across groups in Experiment 2. Black histograms represent the Traditional PDC-HS Group and the white histograms represent the Digital PDC-HS group. The legend applies to both graphs.

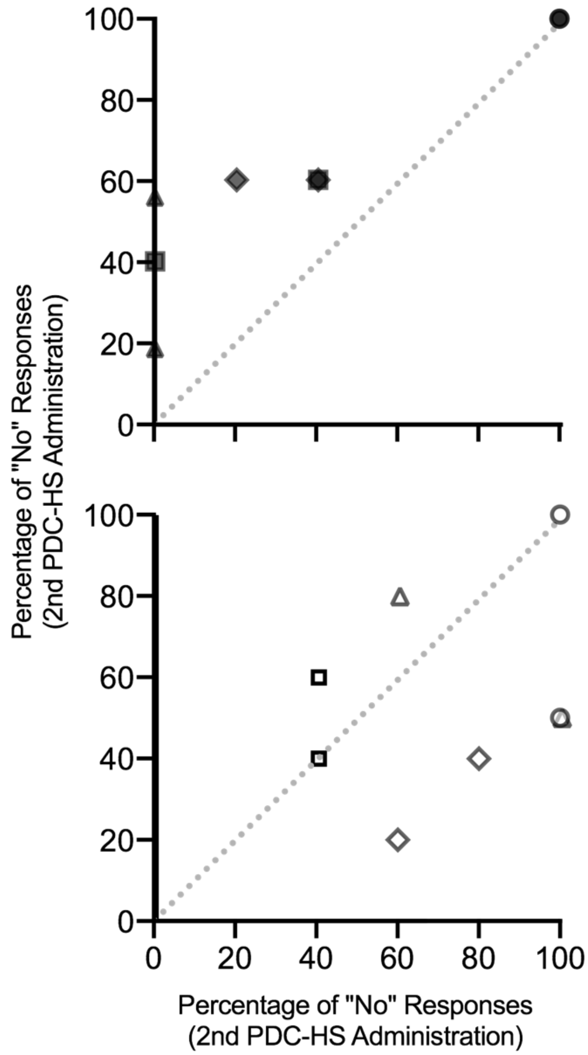

Figure 19.

Correlations between the percentage of “No” responses on the first and second administrations of the PDC-HS across groups in Experiment 2. The Traditional PDC-HS Group is on top and the Digital PDC-HS Group is on the bottom. The darker the data point indicates overlapping data points. The legend applies to all graphs.

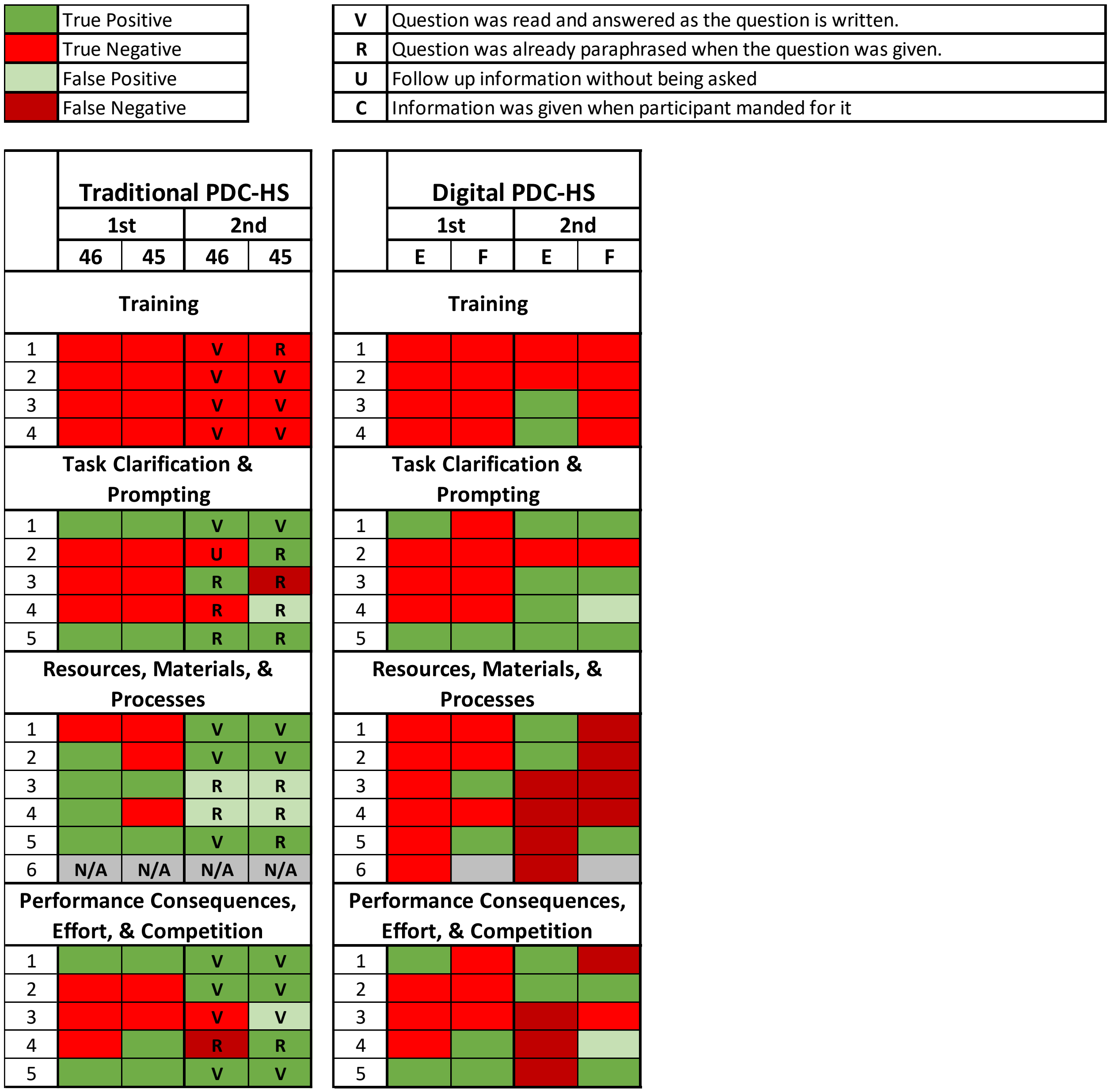

Figure 20.

The SD and Verbal Score Analyses of Experiment 2. The legend applies to both tables. PDC-HS results of Experiment 2 by question while also displaying the SD analysis. Verbal scores were calculated to determine if verbal behavior affected participant responding. Verbal Score definitions for V – Verbatim, R – Rephrase, U – Unprompted, and C – Clarification are in Appendix D.

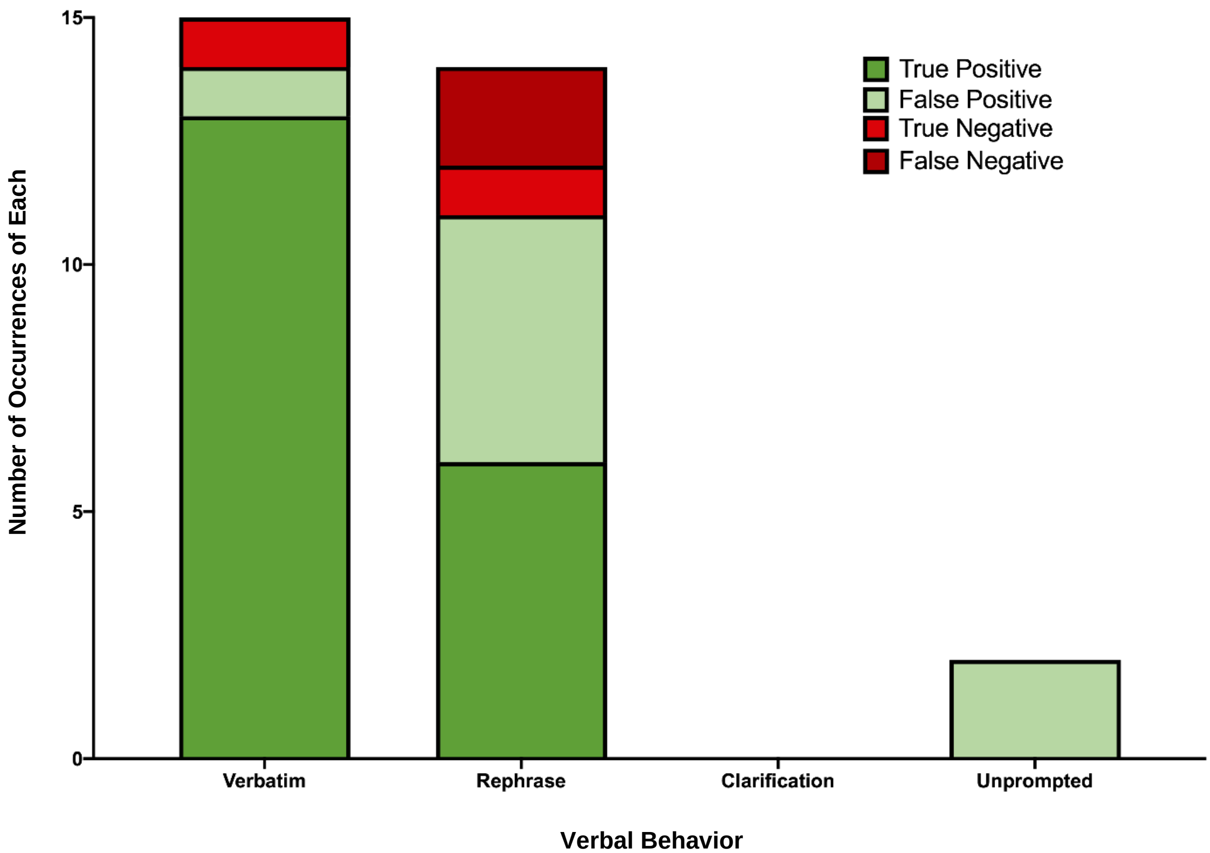

Figure 21.

The verbal behavior of question delivery stratified by the four SD analysis categories for all participants in the Traditional PDC-HS group. Scaled to the *x*-axis is the question delivery (i.e., verbatim, rephrase, clarification, and unprompted; Table 2) and scaled to the *y*-axis is the number of occurrences of each response. Green represents true positives, light green represents false positives, red represents true negatives, and maroon represents false negatives.

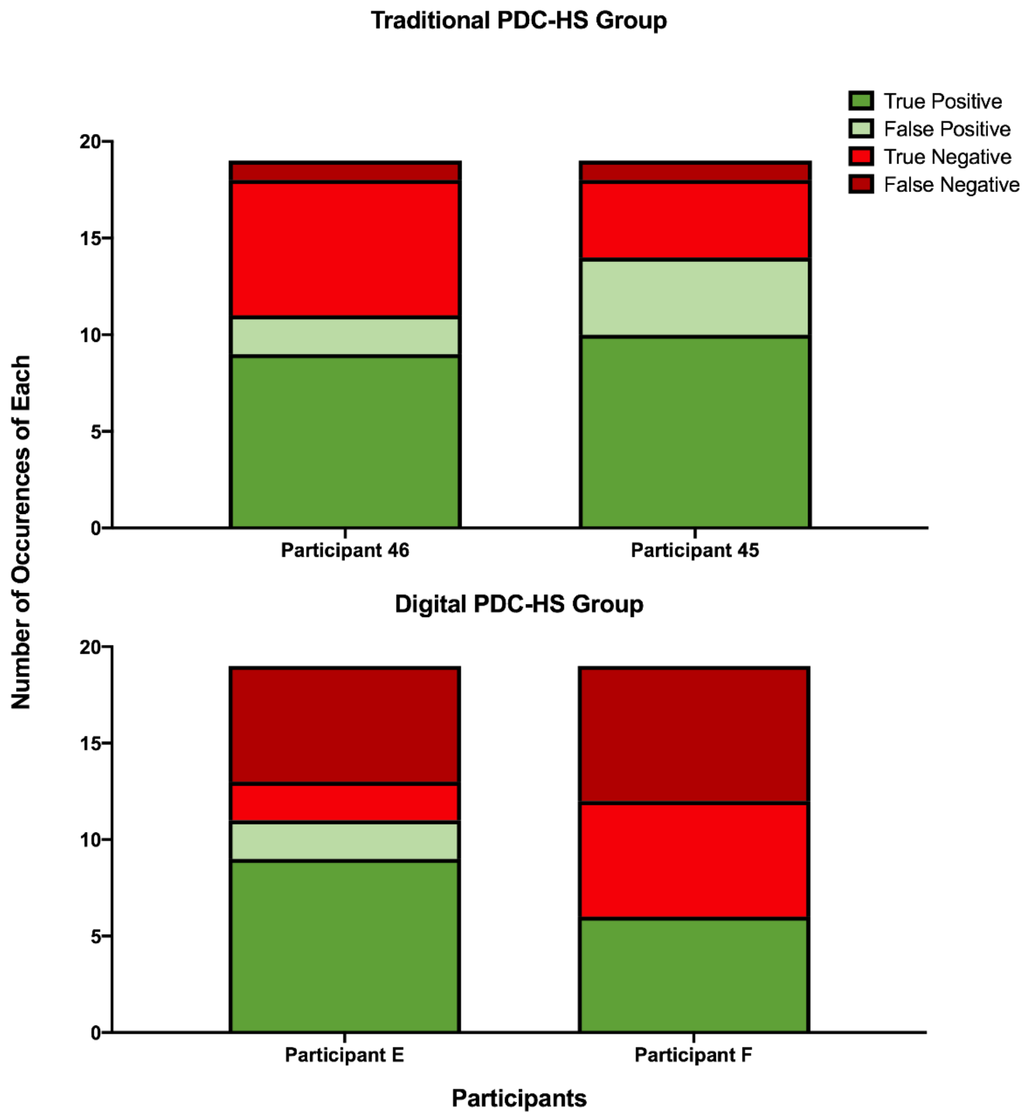

Figure 22.

The graph displays the number of participant responses in each of the four SD categories (i.e., true positive, false positive, true negative, and false negative). Scaled to the *x*-axis are participants and scaled to the *y*-axis is the number of occurrences of each response. Each histogram represents one participant. Green represents true positives, light green represents false positives, red represents true negatives, and maroon represents false negatives (see Table 2). The legend applies to both graphs.

Table 1.

Scoring Guidelines

Question	True Positive	False Positive	True Negative	False Negative
Has the participant received formal training on this task? If yes, check all applicable training methods:	N/A	Yes	No	N/A
❍ Instructions	N/A	Selected	Not Selected	N/A
❍ Demonstration	N/A	Selected	Not Selected	N/A
❍ Rehearsal	N/A	Selected	Not Selected	N/A
Can the participant accurately describe the target task and when it should be performed?	Participant describes task accurately	N/A	Participant describes tasks inaccurately	N/A
Is there evidence that the participant has accurately completed the task in the past?	States yes and participant has solved puzzle	States yes and participant has not solved puzzle	States no and participant has not solved puzzle	States no and participant has solved puzzle
If the task needs to be completed quickly, can the participant perform it at the appropriate speed?*	States yes and participant has solved puzzle in under a minute	States yes and participant has not solved puzzle in under a minute	States no and participant has not solved puzzle in under a minute	States no and participant has solved puzzle in under a minute
Has the participant been informed that he/she is expected to perform the task?	Yes	N/A	N/A	No

Question	True Positive	False Positive	True Negative	False Negative
Can the participant state the purpose of the task?	Participant describes task accurately	N/A	Participant describes tasks inaccurately	N/A
Is a job aid (e.g., a checklist, data sheet) for completing the task visibly located in the task area?	Yes	N/A	N/A	No
Is the participant ever verbally, textually, or electronically reminded to complete the task?	N/A	Yes	No	N/A
Is the task being performed in an environment well-suited for task completion (e.g., not noisy or crowded)?	Yes	N/A	N/A	No
Are there sufficient numbers of trained staff available to complete the task?	Yes	N/A	N/A	No
If materials (e.g., teaching stimuli, preferred items) are required for task completion, are they readily available (e.g., easy to find, nearby)? If no materials are required, proceed to question 5.	Yes	N/A	N/A	No

Question	True Positive	False Positive	True Negative	False Negative
List Materials Below	N/A	Selected	N/A	N/A
Listed Materials	Puzzle/Directions	Any other materials	Incorrect Materials NOT listed or left blank	Puzzle or Written Instructions not listed (results in one or two FNs) only when space is available
Are the materials necessary to complete the task well designed for their intended purpose?	Yes	Did not accurately list materials in previous step (must list only and all correct materials)	N/A	No
Are the materials necessary to complete the task well organized for their intended purpose?	Yes	Did not accurately list materials in previous step (must list only and all correct materials)	N/A	No
Can the task be completed without first completing other tasks? If not, indicate below the tasks that must be completed first.	Yes	N/A	N/A	No
Tasks	N/A	Any tasks listed	Question is skipped due to previous answer	Question is presented due to previous answer but skipped

Question	True Positive	False Positive	True Negative	False Negative
If you answered NO for Question 5, are other participants responsible for completing any of the earlier tasks in the process? If so, indicate the participant(s) below.	N/A	Yes	Question is skipped and answered yes previous question	No
Tasks	N/A	Names listed	Question is skipped and answered yes previous question	Previous question answered "No" and "skipped"
Is the participant ever directly monitored by a supervisor?	Yes	N/A	N/A	No
Frequency	Continuously	Any other options selected	N/A	Not answered due to question being skipped
Does the participant ever receive feedback about the performance?	Yes	N/A	N/A	No
By whom?	Supervisor/Meagan	Any other person	N/A	No one selected
How Often?	After every trial	Other duration	N/A	Question is skipped due to previous answer
Delay from task?	Immediately	Other time given	N/A	Question is skipped due to previous answer

Question	True Positive	False Positive	True Negative	False Negative
Feedback Focus:	Positive	Corrective	N/A	Question is skipped due to previous answer
Feedback Type:	Verbal	Other options selected	N/A	Question is skipped due to previous answer
Does the participant ever see the effects of accurate task completion?	Yes if participant has solved it	Yes if participant has not solved it	No if participant has not solved it	No if participant has solved it
How?	Puzzle is completed/ Point is earned	Answer is given if participant has not solved it, or wrong answer	Question is skipped due to participant not solving the puzzle	Question is skipped due to previous answer but should have been answered
Is the task simple or does it involve relatively low response effort?	Yes	N/A	N/A	No
Does the task generally take precedence over other potentially competing tasks?	Yes	N/A	N/A	No
List Tasks Below				
Tasks	N/A	Task given	Question skipped due to previous answer or not answered	Question presented due to previous answer but not answered

Table 2.

Operational Definitions

Dependent Variable	Operational Definition
Duration	The amount of times in s that it takes to solve the puzzle (up to two minutes). *Onset*: Verbal prompt of "go" or "you may begin," etc. *Offset*: 9th tile replaces gray square or two minutes have elapsed *Example*: 96 s *Non Example*: 240 s
Actual Number of Moves	The number of moves that are completed within a trial. A click that results in a tile moving. If a participant clicks a tile, and then moves it back to the original location, it is counted as two moves.
Minimum Number of Moves	The minimum number of moves it would take to solve the puzzle based on the starting configuration. All puzzle configuration can be solved ranging from 1 to 31 moves. *Example*: 26 moves *Non Example*: 45 moves
Duration of Behavioral Skills Training (BST)	The amount of time it took to complete BST, measured in minutes and s. *Onset*: Pressing play on the initial training video *Offset*: After feedback is provided on the timed practice trial *Example*: 22 min and 18 s *Non Example*: 783 s

Sensitivity	This measure calculated the ability of the PDC-HS to accurately identify the components of task completion that were *available* or the amount of "Yes's" on the PDC-HS. Sensitivity was calculated by dividing the number of correct hits by the sum of correct hits and misses.
Specificity	This measure calculated the ability of the PDC-HS to accurately identify components of task completion that were *missing* or the number of "No's" on the PDC-HS. Specificity was calculated by dividing the number of correct rejections by the sum of correct rejections and false alarms.
Signal Detection	Signal Detection was completed evaluating four measures: *True Positives*: a participant accurately identifying that an element of task completion had taken place (e.g., indicating that they had solved the puzzle when they had). *False Positives*: Inaccurately stating that an element of task completion was implemented when it had not been (e.g., indicating that they had received training). *True Negative*: accurately stating that a specific element of task completion had not taken place (e.g., they had not received training) *False Negatives*: neglecting to identify an element of task completion (e.g., indicating that they had not been told to solve the puzzle even though they had)

Verbal Score	This measure calculates the amount of the Traditional PDC-HS that was read verbatim: <u>Verbatim</u>: Question was read and answered as the question is written. *Example*: "Did you receive formal training on how to solve the puzzle quickly?" *Non Example*: "Did you get trained on how to solve the puzzle?" <u>Reworded</u>: Question was already paraphrased when the question was given. *Example*: First time asking question: "Did you receive training on how to solve the puzzle?" *Non Example*: Reading written question and then changing it to example <u>Unprompted</u>: Follow up information without being asked *Example*: Reading verbatim example and reworded example back to back without any response from participant. *Non Example*: Participant is read question, waits, and then the investigator provides clarification. <u>Clarification</u>: Information was given when participant manded for it *Example*: Question is read and participant says "Can you phrase that differently?" and more information is offered *Non Example*: Question is read and follow up information is immediately given

Table 3.

Procedural Fidelity Checklist

Baseline	**Complete**
Puzzle corresponds with correct trial (Puzzle #3 for trial #3)	
No other tabs are open in participant's browser	
Confirm that the participant can see the entire screen	
Entire puzzle should show on the screen	
Investigator camera should be off	
Investigator stops participant by 2 minutes	
Domain Manipulation	
Puzzle corresponds with correct trial (Puzzle #3 for trial #3)	
Only Puzzle, Background Information, and Written instructions are open in the *same* browser when trial begins	
Only Puzzle and Checklist are open (Checklist may be downloaded)	
Confirm that the participant can see the entire screen	
Entire puzzle should show on participant's screen	
Investigator camera should be on	
Investigator stops participant by 2 minutes	
Accurate feedback is provided	
Barrier Removal	
Puzzle corresponds with correct trial (Puzzle #3 for trial #3)	
Only Puzzle, Background Information, and Written instructions are open in the *same* browser when trial begins	
Only Puzzle and Checklist are open (Checklist may be downloaded)	
Confirm that the participant can see the entire screen	
Entire puzzle should show on the screen	
Investigator camera should be on	
Investigator stops participant by 2 minutes	
Feedback is provided	

Assessments of the Digital and Traditional PDC-HS

PDC-HS Type	Sensitivity	Specificity	PPV	NPV
Traditional	62%	77%	71%	71%
Digital	51%	61%	56%	56%

Table 4.

Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) for Experiment 1.

Assessments of the Digital and Traditional PDC-HS

PDC-HS Style	Sensitivity	Specificity	PPV	NPV
Traditional	88%	78%	70%	92%
Digital	56%	75%	78%	58%

Table 5.

Sensitivity, Specificity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV) for Experiment 2.

Appendix A

Sliding Puzzle Consent Script

Key Information:

- This project is studying barriers to fluent task completion.
- Your participation in this research project is completely voluntary.
- Your participation will take about one hour.
- You will be asked to do the following procedures: solve a 3x3 sliding puzzle multiple times, take a survey, and solve a 3x3 sliding puzzle multiple times again. More detailed information on the procedures can be found below. This process may be completed multiple times.
- There are no possible risks or discomforts related to the study.
- This study will benefit others by creating a digital system to help employees at work.
- You may choose not to participate or may participate in an alternate study.

As a student in the University of Kansas's Department of Applied Behavioral Analysis, I am conducting a research project about quick task completion. If you decide to participate in the study, you will be asked to complete a task (a 3x3 sliding puzzle) multiple times, complete a survey, followed by the puzzle multiple times. You may be asked to repeat the process more than once. Your participation is expected to take about 60 minutes. You have no obligation to participate and you may discontinue your involvement at any time. If you choose to end early, any data that are already collected on you may be used for future publications.

Your participation should cause no more discomfort than you would experience in your everyday life. Although completing may not benefit you directly, you may receive extra credit for participating (at the discretion of your instructor) and the information obtained from the study will help us gain a better understanding of assessments used to determine barriers to task completion. Your identifiable information will not be shared unless (a) it is required by law or university policy, or (b) you give written permission.

This study is being recorded. Recording is required to participate. Only I, graduate students in my lab, and my faculty supervisor will have access to recordings which will be stored on a secure server and will be destroyed upon thesis defense, study publication, or after seven years.

Participation in the study indicates your willingness to take part in this study and that you are at least 18 years old. Should you have any questions about this project or your participation in it you may ask me (or contact me at meaganwylie@ku.edu or my faculty supervisor, Robin Kuhn, in the Department Applied Behavioral Analysis, at rmkuhn@ku.edu. If you have any questions about your rights as a research participant, you may call the Human Research Protection Program at (785) 864-7429 or email irb@ku.edu.

Do you have any questions?

Please state your name, your age, and if you consent to participate.

Please also state your major, year in school, and your GPA if you feel comfortable.

Appendix B

Study Procedures

1. Hello, you will be completing sliding puzzles today. Your goal is to solve the puzzle. You'll know when it is complete when the ninth tile replaces the gray one and completes the picture. The numbers correspond to the numerical order of the pieces. To help me score your accuracy, I will need to see your screen as you are solving the puzzle. I will be sending you links to the puzzle via Zoom chat. Each trial will be a new link. Nothing else can be open in your browser during this time.

2. Traditional (Interview) vs. Digital (Survey)
    a. When you are interviewed, I will type click or type in the answers.
    b. When you take the survey, I will chat it in the link to you and you will open it up on your screen. I will leave the room for five minutes to give your privacy.

3. My camera will be on and off throughout the study.

Appendix C

Task: Solving the Sliding Puzzle
Steps to Solve the Puzzle

Place tiles 1 and 3 in the positions below.
Place tile #2 under tile #3.

1	3	4
5	2	6
8	7	

1	2	3
5		4
8	7	6

Tiles 7 and 4 need to be placed in the positions below.

1	2	3
7	4	
8	5	6

Tile 7 now needs to be moved down to the bottom row on and the other tiles can be moved into place.

1	2	3
4		6
7	8	5

Then move the tiles 5, 6, 7 into the appropriate spots.

1	2	3
4	5	6
7	8	

Appendix D

In this survey you will answer questions about your participation in the study session so far and about the puzzle task.

Did you receive formal training on how to solve the puzzle quickly?

○ Yes

○ No

Check all applicable training methods. (Select all the apply)

☐ Instructions

☐ Demonstration

☐ Rehearsal

Can you describe the target task?

____

Have you accurately completed the puzzle anytime during this session?

○ Yes

○ No

Have you been able to solve the puzzle at the appropriate speed?

○ Yes

○ No

Have you been informed that you are expected to solve the puzzle?

- ◯ Yes
- ◯ No

What is the purpose of this task?

____

Is a job aid (e.g., a checklist, data sheet, written instructions) for completing the task visibly located in the task area?

- ◯ Yes
- ◯ No

Is the employee ever verbally, textually, or electronically reminded to complete the task?

- ◯ Yes
- ◯ No

Are you solving the puzzle in an environment well suited for task completion (e.g., not noisy or crowded)?

- ◯ Yes
- ◯ No

Are there sufficient number of people trained to complete the puzzle?

- ◯ Yes
- ◯ No

Are materials that are required for solving the puzzle readily available?

- ◯ Yes
- ◯ No
- ◯ Not Applicable

List materials required for solving the sliding puzzle.

	Is it available?	
	Yes	No
Item 1	◯	◯
Item 2	◯	◯
Item 3	◯	◯
Item 4	◯	◯

Are the materials necessary to solve the puzzle quickly **well designed** for their intended purpose?

- ◯ Yes
- ◯ No

Are the materials necessary to solve the puzzle quickly **well organized** for their intended purpose?

- ◯ Yes
- ◯ No

Can the task be completed without first completing other tasks?

- ◯ Yes
- ◯ No

Which tasks need to be completed first?

- [ ] Task 1 ______________________________
- [ ] Task 2 ______________________________
- [ ] Task 3 ______________________________
- [ ] Task 4 ______________________________

Are other staff members responsible for completing any of the earlier tasks in the process?

- ○ Yes
- ○ No

Indicate the employee(s) below.

- [ ] Task 1 ______________________________
- [ ] Task 2 ______________________________
- [ ] Task 3 ______________________________
- [ ] Task 4 ______________________________

Are you ever directly monitored by a supervisor while solving the puzzle?

- ○ Yes
- ○ No

Indicate the frequency of supervisor monitoring. (Select all that apply)

- ☐ Hourly
- ☐ Daily
- ☐ Weekly
- ☐ Monthly
- ☐ Other ____

Do you ever receive feedback about your performance?

- ○ Yes
- ○ No

Indicate below.

- ☐ By whom? ____
- ☐ How often? ____
- ☐ How soon after the puzzle is completed?

____

What is the focus of your feedback? (Select all that apply)

- ☐ Positive
- ☐ Corrective

How is feedback given? (Select all that apply)

- ☐ Written
- ☐ Verbally
- ☐ Graphed
- ☐ Other ____

Do you ever see the effects of accurate task completion?

- ○ Yes
- ○ No

How?

____

Is the task simple or does it involve relatively low effort?

- ○ Yes
- ○ No

Does solving the puzzle generally take precedence over other potentially competing tasks?

- ○ Yes
- ○ No

Indicate these tasks below.

- ☐ Task 1 ____
- ☐ Task 2 ____
- ☐ Task 3 ____
- ☐ Task 4 ____

Is there anything else that should be known in regards to solving the puzzle quickly?

____

Appendix E

Baseline Script

“Complete the puzzle. This trial will end after you solve the puzzle, until you run out of time, or until you choose to stop the trial. I cannot answer any questions while you are solving the puzzle or in between sessions. You may begin when I tell you to start.”

Appendix F

PDC-HS Script

Traditional PDC-HS

“You are now going to have an opportunity to complete an interview about your experience. This interview should take about five minutes. Your responses will not be analyzed until your participation has ended. Your answers will not affect your extra credit.”

Digital PDC-HS

“You are now going to have an opportunity to complete a survey about your experience. I will leave for five minutes to allow you to complete the survey and then I will come back. Your survey responses will not be reviewed until your participation has ended. I cannot answer any questions about this survey or its wording.”

Appendix G

Background Information

“You are currently working in an office that sells paper. Your office has a friendly pool to pick winners of Major League Baseball games. However, due to recent health concerns, opening day has been indefinitely postponed and your business is looking for other ways to entertain themselves.”

“The office decides on having a puzzle competition where you are to solve a sliding puzzle. You have written instructions to help you. Everyone in your office has already been trained to solve the puzzle. Your supervisor, Meagan, will be watching to ensure no one cheats. Start the puzzle when you hear “go”.”

Appendix H

The domain manipulations included in the present study during baseline and domain manipulations conditions. Red writing denotes items expected to be identified as a barrier. Green writing denotes items not expected not be identified as a barrier. Blue writing denotes items that may or may not be identified as a barrier depending upon the participant's performance in the study.

Performance Diagnostic Checklist – Human Services

TRAINING

1	❍ Yes ❍ No	Has the participant received formal training on this task? If yes, check all applicable training methods: ❍ Instructions ❍ Demonstration ❍ Rehearsal Participants did not receive any training on how to solve the puzzle quickly.
2*	❍ Yes ❍ No	Can the participant accurately describe the target task and when it should be performed?* Participants were never be told the target task (i.e., to complete the puzzle *in under* a minute.)
3	❍ Yes ❍ No	Is there evidence that the participant has accurately completed the task in the past? To participate in this study, participants must not have solved the puzzle three times consecutively in under a minute during baseline, however participants may have solved the puzzle in under a minute.
4*	❍ Yes ❍ No ❍ N/A	If the task needs to be completed quickly, can the participant perform it at the appropriate speed?* To participate in this study, participants must not have solved the puzzle three times consecutively in under a minute during baseline, however participants may have solved the puzzle in under a minute.

TASK CLARIFICATION & PROMPTING

1	❍ Yes ❍ No	Has the participant been informed that he/she is expected to perform the task? Participants were told that they are expected to solve the puzzle.

2*	❍ Yes ❍ No	Can the participant state the purpose of the task? Participants were told that they need to solve the puzzle to win a point in the company's puzzle competition.
3*	❍ Yes ❍ No	Is a job aid (e.g., a checklist, data sheet) for completing the task visibly located in the task area? Participants were provided with a job aid.
4	❍ Yes ❍ No	Is the participant ever verbally, textually, or electronically reminded to complete the task? Participants were never be reminded to complete the task.
5	❍ Yes ❍ No	Is the task being performed in an environment well-suited for task completion (e.g., not noisy or crowded)? Participants solved the puzzle in a quiet environment.

RESOURCES, MATERIALS, & PROCESSES

1	❍ Yes ❍ No	Are there sufficient numbers of trained staff available to complete the task? Participants were told that there are enough people in their office have been taught how to solve the puzzle.
2*	❍ Yes ❍ No ❍ N/A	If materials (e.g., teaching stimuli, preferred items) are required for task completion, are they readily available (e.g., easy to find, nearby)? If no materials are required, proceed to question 5. List materials below and indicate their availability. Item 1: ____________________________ Item 2: ____________________________ Item 3: ____________________________ Item 4: ____________________________ Materials were made available to participants.
3*	❍ Yes ❍ No ❍ N/A	Are the materials necessary to complete the task well designed for their intended purpose? Materials were well designed.
4*	❍ Yes ❍ No ❍ N/A	Are the materials necessary to complete the task well organized for their intended purpose? Materials were well organized.
5	❍ Yes ❍ No	Can the task be completed without first completing other tasks? If not, indicate below the tasks that must be completed first.

		Task 1: ___________________________ Task 2: ___________________________ Task 3: ___________________________ Task 4: ___________________________ No other tasks need to be completed first.
6	❍ Yes ❍ No ❍ N/A	If you answered NO for Question 5, are other participants responsible for completing any of the earlier tasks in the process? If so, indicate the participant(s) below. Task 1: ___________________________ Task 2: ___________________________ Task 3: ___________________________ Task 4: ___________________________ Question was asked or skipped depending upon the answer to number 5.

PERFORMANCE CONSEQUENCES, EFFORT, & COMPETITION

1	❍ Yes ❍ No	Is the participant ever directly monitored by a supervisor? If so, indicate the frequency of monitoring. ❍ hourly ❍ daily ❍ weekly ❍ monthly ❍ Other: ____________________ Participants were observed during this trial.
2	❍ Yes ❍ No	Does the participant ever receive feedback about the performance? If yes, indicate below. By whom? Meagan/Supervisor How often? After every trial Delay from task? Immediately Check all that apply: Feedback Focus: ❍ Positive ❍ Corrective Feedback Type: ❍ Written ❍ Verbal ❍ Graphed ❍ Other: Participants received feedback on their solving speed and if they scored a point in the puzzle competition.
3	❍ Yes ❍ No	Does the participant ever see the effects of accurate task completion? If yes, how? This answer is dependent on if the puzzle was solved.
4	❍ Yes ❍ No	Is the task simple or does it involve relatively low response effort? The puzzle is a simple task that requires low response effort.

5	❍ Yes ❍ No	Does the task generally take precedence over other potentially competing tasks? If not, indicate these competing tasks below. Task 1: ___________________________ Task 2: ___________________________ Task 3: ___________________________ Task 4: ___________________________ There were no competing tasks.

Appendix I

BST Procedures

1. Restate the goal: The goal of this task is to solve the puzzle in under a minute.

2. Gray Square: Create space using the gray square: when solving this puzzle, an important element to consider is where the gray square is. The gray square creates an opening that you can slide a tile into. Therefore, the gray square should precede where the tile is going.

3. Place tiles 1-3
   a. State the step: The first step that you need to complete is to place tiles 1, 2, and 3 in the top row. One way to make this step easier is to have tile 1 in the top left corner, tile 3 in the top middle and tile 2 directly under tile 3. Then, you will move the tile that is in the top right corner, move tile three over, and slide tile two into the top middle space.
   b. Model twice
      i. There is a tricky combination that I need to show you how to solve. I am going to intentionally put the tiles in that position, and I will show you how to solve it.
      ii. If tiles 3 and two are inverted, you have to separate tiles one and three to disrupt the pattern. After tiles one and three are separated, you bring the two to the bottom row and attempt to place tiles 1, 2, and 3 into the correct location again.
      iii. I am going to reset and solve it one more time and then you can take a turn.
   c. Participant's turn
   d. Provide feedback: Feedback will vary depending on participant performance.
   e. Participants must do this step correctly three times

4. Place tiles 7 and 4
   a. State the step: I am going to teach you how to do the first and second steps now.
   b. Model twice
      i. Provide insight when 4/7 are in the wrong order
      ii. If tiles seven and four are inverted (i.e., tile four is where tile seven should be and vice versa), separate the two tiles and bring them together again. Given that there are so few tiles to move at this stage, it is extremely difficult to only move one tile without disrupting the others. Therefore, your time is better spent separating the tiles and bringing them together at a later step.
   c. Participant's turn
   d. Provide feedback: Feedback will vary depending on participant performance.
   e. Participants must do this step correctly three times
   f. Put tiles 1, 2, 3, 7, 4 in the correct location.

5. Place tiles 5, 6, 8 Now I am going to teach you all three steps combined. This means tiles 1, 2, and 3 will be at the top, tiles 7 and 4 will be in the middle row, and then tiles 5, 6, 8 and will be slid into place.
    a. State the step: Since there are only three tiles for this stage, you simply rearrange the tiles in the order that you need them. You will know that you are done when tile nine replaces the gray square.
    b. Model twice
    c. Participant’s turn
    d. Provide feedback: Feedback will vary depending on participant performance.
    e. Participants must do this correctly three times

6. Solve entire puzzle in under a minute
    a. Must do this final step at least once completely independently
    b. Can continue to practice

Appendix J

Debriefing Statement

This study included four participant groups. Each participant group received instructions and procedures with a different part omitted at the beginning of the study. After completing the survey, the part that was initially omitted was added in to help you solve the puzzle. For example, participants in your group did not receive complete training on how to solve the puzzle until after you filled out the survey about the puzzle. The instructions and procedures were omitted to see if participants would accurately identify the barrier(s) to quick puzzle completion (in your case training). If participants accurately identify the puzzle completion barriers on the survey, then this procedure can be used to compare the results of a digital version of the survey to the typical interview-style version of the survey. The results of the study will be evaluated recognizing that incomplete disclosure may have altered participant motivation to complete the task. You may withdraw from the study at this time. You may also withdraw your consent for the use of the audio or video recordings; however, this would withdraw you from the study. Thank you for taking part in the study.

Appendix K

Video Transcript:

Okay let’s add some background information and figure out why you are solving this puzzle. You are currently in an office that sells paper. Your office usually has a pool to pick winners of major league baseball games but due to health concerns, the season has been cancelled. Bummer. Instead, your office decides to have a sliding puzzle competition. Game on! There are enough people trained on how to complete this puzzle because everyone has been taught how to solve it, except for you. Bummer number 2. The only materials you will need are the puzzle and the checklist. Your supervisor, Meagan, will send those materials to you. So, you are probably thinking what is the purpose of solving the puzzle? You need to solve the puzzle as fast as possible to win the competition. Your supervisor is really into this competition. If your phone rings or somebody asks you a question, you don’t have to answer. Nothing else takes precedence over this task. You can solve the puzzle without any distractions. One last important thing, your supervisor Meagan will be continuously supervising you and giving you feedback after every puzzle. When you take the survey, try to answer the questions using this video to help you. Happy puzzle solving! Remember, to get points in the puzzle competition, you have to solve the puzzle. Have fun!

Appendix L

Puzzle Checklist

Use this list to help you track how many puzzles you have solved.

Puzzle 1

- ◯YES
- ◯NO

Puzzle 2

- ◯YES
- ◯NO

Puzzle 3

- ◯YES
- ◯NO

Puzzle 4

- ◯YES
- ◯NO

Puzzle 5

- ◯YES
- ◯NO

Puzzle 6

- ◯YES
- ◯NO

Puzzle 7

- ◯YES
- ◯NO

Puzzle 8

- ◯YES
- ◯NO

Puzzle 9

◯YES
◯NO

Puzzle 10

◯YES
◯NO

Puzzle 11

◯YES
◯NO

Puzzle 12

◯YES
◯NO

Puzzle 13

◯YES
◯NO

Puzzle 14

◯YES
◯NO

Puzzle 15

◯YES
◯NO

Puzzle 16

◯YES
◯NO

Puzzle 17

◯YES
◯NO

Puzzle 18

◯YES
◯NO

Puzzle 19

◯YES
◯NO

Puzzle 20

◯YES
◯NO

Puzzle 21

◯YES
◯NO

Puzzle 22

◯YES
◯NO

Puzzle 23

◯YES
◯NO

Puzzle 24

◯YES
◯NO

Puzzle 25

◯YES
◯NO