Assessing the Validity and Reliability of the Performance Diagnostic Checklist-Human Services 1.1

Sebastian Jimenez
Assessing the Validity and Reliability of the Performance Diagnostic Checklist-
Human Services 1.1

by

Sebastian Jimenez

A thesis submitted to the School of Behavior Analysis of
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Applied Behavior Analysis and Organizational Behavior Management

Melbourne, Florida
July 2022
We the undersigned committee hereby approve the attached thesis, “Assessing the Validity and Reliability of the Performance Diagnostic Checklist-Human Services 1.1” by Sebastian Jimenez.

David A. Wilder, Ph.D., BCBA-D
Professor
School of Behavior Analysis
Major Advisor

Patrick D. Converse, Ph.D.
Professor
School of Psychology

Rachael E. Ferguson, Ph.D.
Assistant Professor
School of Behavior Analysis

Robert A. Taylor, Ph.D.
Professor and Dean
College of Psychology and Liberal Arts
Abstract

Title: Assessing the Validity and Reliability of the Performance Diagnostic Checklist-Human Services 1.1

Author: Sebastian Jimenez

Advisor: David A. Wilder, Ph.D., BCBA-D

The PDC-HS is an informant-based assessment tool used by OBM practitioners in human service settings. An updated version known as the PDC-HS 1.1 has recently been released. The present study measured the validity and reliability of the PDC-HS 1.1 by analyzing answers obtained while watching a video of a simulated interview between a consultant and a supervisor. Three video vignettes were created, each describing a performance concern in one or more areas of the PDC-HS 1.1. Twenty-one participants watched all vignettes and filled out the PDC-HS 1.1 based on the answers provided. Validity was measured by calculating the percentage of participants who correctly identified the area(s) in the PDC-HS 1.1 responsible for the performance concern presented in each vignette. Participants repeated the assessment about two weeks later to assess test-retest reliability. Interrater reliability was measured by pairing participants randomly and comparing their scores. In addition, an intervention selection component was included to assess whether a corresponding intervention was selected to target the indicated domain. Results demonstrate that about 90% of participants correctly identified the indicated area of the PDC-HS 1.1 and 79% selected an appropriate intervention. Test-retest and interrater reliability scores were above 85% and demonstrate that...
the tool is generally reliable. The results provide support for the use of informant-based assessments in human services settings and suggest that participants with relatively little experience in behavior analysis can conduct assessment interviews accurately and reliably.

*Keywords:* Interrater reliability; intervention selection; Performance Diagnostic Checklist-Human Services 1.1; test-retest reliability; validity
# Table of Contents

Table of Contents ..................................................................................................... v
List of Tables .......................................................................................................... vi
Acknowledgment ................................................................................................... vii

Chapter 1 Assessing the Validity and Reliability of the Performance Diagnostic Checklist-Human Services 1.1.............................................................. 1

Chapter 2 Method .................................................................................................. 21
  Video Vignettes.............................................................................................................. 22
  Data Collection and Dependent Variables ............................................................... 24

Chapter 3 Results ................................................................................................... 27
  Validity........................................................................................................................... 27
  Test-retest reliability ..................................................................................................... 29
  Interrater reliability...................................................................................................... 30

Chapter 4 Discussion ............................................................................................. 32

References ............................................................................................................... 43

Appendix ................................................................................................................. 50
List of Tables

Table 1 - PDC-HS 1.1 Scoring Summary ................................................................. 28
Table 2 - Summary of PDC-HS 1.1 Test-retest Reliability ................................. 30
Table 3 - Summary of PDC-HS 1.1 Interrater Reliability .................................... 30
Table 4 - Summary of Intervention-Selection Accuracy ...................................... 31
Acknowledgment

Thank you to my supervisor, David Wilder, for their continued support and guidance throughout this process. Thank you to my committee members for providing their advice and expertise. Thank you to Victoria German and Megan Galban for their invaluable assistance preparing research materials.
Chapter 1
Assessing the Validity and Reliability of the Performance Diagnostic Checklist-Human Services

1.1

Behavior analysis is a natural science that comprises four interrelated domains: (1) philosophy; (2) basic research; (3) applied research; and (4) practice (Cooper et al., 2019). Behaviorism is the philosophy of behavior analysis and posits that observable behavior should be the independent variable in psychology, not states of mind or mental processes. Most of the basic research in behavior analysis has been done in the experimental analysis of behavior (EAB), whereas much of the effort for developing technologies for improving behavior of social significance has been done in applied behavior analysis (ABA). Behavior analysis in practice involves applying the concepts and principles of behavior analysis in professional practice. This paper is focused on the domain of applied research which is conducted primarily through ABA.

The field of ABA is considered to have begun with the inception of the *Journal of Applied Behavior Analysis* in 1968. Seven dimensions of ABA were outlined by Baer, Wolf, and Risley (1968), who proposed that applied research in
behavior analysis be focused on matters of social importance. These seven dimensions were: (1) applied; (2) behavioral; (3) analytic; (4) technological; (5) conceptually systematic; (6) effective; and (7) generality. The goal of applied behavior analysis has been to make socially significant improvements in certain behaviors deemed severe or problematic while analyzing possible environmental factors responsible for the improvements (Cooper et al., 2019).

A critical component of ABA has been the use of functional assessments to identify maintaining variables for targeted behaviors (Austin, 1999). Functional assessments analyze the variables surrounding behavior to identify the functions of (often) problematic behavior. Using functional assessments is viewed as ‘best practice’ in clinical settings and has allowed behavior analysts to identify functional relationships between behavior and the environment. Once functional relationships are identified, interventions are developed to target maintaining variables. These functionally relevant interventions have generally consisted of two interrelated strategies: (1) the contingency between problem behavior and maintaining consequence is weakened, and (2) a contingency between a replacement behavior and the maintaining consequence for the problem behavior is established or strengthened (Mace, 1994). Often, the consequences which have maintained the problem behavior are programmed to maintain appropriate behavior.

Three general approaches to functional assessments exist: informant assessments, descriptive assessments, and experimental analysis (Lennox &
Informant assessments refer to the collection of information via indirect means such as interviews, questionnaires, surveys, or rating scales. These assessments take relatively little time to complete and are simple to administer. There have been several informant assessments developed in the clinical behavior-analytic literature including: the motivation assessment scale (MAS; Durand & Crimmins, 1992); the functional analysis screening tool (FAST; Iwata et al., 2013); and questions about behavioral function (QABF; Paclawskyj et al., 2000). However, the information obtained through these assessments has limited value due to a lack of direct observation or experimental manipulation (Austin, 1999). Informant assessments have been used in conjunction with descriptive or experimental techniques to provide preliminary information which guides further assessments.

Descriptive assessments are conducted through direct observation of behavior in natural settings and require a more rigorous approach (Austin, 1999). The most common descriptive assessment methods have been the ABC analysis and scatterplot recording (Cooper et al., 2019). Through direct observation, behavior analysts have been able to identify and describe environmental events that occur shortly before and after instances of behavior. The information provided by descriptive assessments is correlational rather than causal (Hanley, 2012). It is only through experimental manipulation that these environmental factors can be confidently identified as functions of behavior.
The most researched method of functional assessment is the experimental analysis or functional analysis, wherein relevant environmental variables are directly manipulated while their effects on problem behavior are observed and measured (Cooper et al., 2019). The functional analysis (FA) was introduced by Iwata et al. (1982) as a method of determining the maintaining variables for self-injurious behavior. The basic procedure involves various conditions, based on operant function, where only one variable is manipulated at a time and its effect on behavior is recorded. These variables can be found in the individual’s natural environment and are suspected to contribute to the reinforcement and maintenance of problem behavior. For example, in the demand condition of the FA, a task that has previously been reported to be aversive is presented to the participant. If the participant engages in high rates of problem behavior during the demand condition, this suggests the behavior is maintained by escape. The functional analysis methodology has generated an expansive database of extensions and replications across different client populations, target behaviors, and settings (Hanley et al., 2003). Because of this, the FA is widely considered the ‘gold standard’ in identifying the operant function of behavior (Wilder et al., 2019).

The use of functional assessments has extended into organizational settings. Organizational Behavior Management (OBM) is a sub-branch of ABA that applies behavior analytic principles and procedures to improve performance in the workplace using performance analysis (Reid & Parsons, 2000). Conventionally,
performance analysts in OBM have worked at three levels of analysis: the individual level, the process level, and the systems level (Austin, 2000). Performance management focuses on the individual level of analysis and attempts to make improvements in the behavior of individual performers (Daniels & Bailey, 1989). The use of pre-intervention assessments has guided OBM practitioners in identifying variables that contribute to deficits in performance (Austin, 1999). Both indirect (Gilbert, 1978; Mager & Pipe, 1997; Rummler & Brache, 1995) and descriptive assessments (Daniels & Bailey, 1989) have been developed and used in various organizational settings. These assessments have focused on operationalizing skill deficits, describing antecedents and consequences, and identifying appropriate interventions. However, formal experimental analysis is scarce and has not been frequently reported in the OBM literature (Austin, 1999). The reason for this is that experimental analysis requires rigorous control of environment variables which can result in a lengthy and costly process that makes them impractical in organizational settings (Wilder et al., 2020).

Austin (1999) called for an increase in the use of formal functional assessments in OBM and noted three possible reasons for their scarcity in the empirical literature. The first reason was that interventions done in OBM have had relatively high success without the use of formal functional analysis procedures. Behavior analysts implementing interventions in organizational settings have often been able to produce substantial and lasting behavior change. The second reason was
that behavior that occurred in organizational settings was mainly rule-governed. The contingencies controlling rule-governed behavior are not readily apparent while the behavior is occurring, this makes observing environmental factors controlling behavior a difficult if not impossible task. Finally, much of the functional analysis methodology has been focused on decreasing behavior whereas the focus of many OBM interventions has been to increase and maintain desired behavioral repertoires. These reasons, coupled with factors like market competition and the potential costs associated with analog functional analysis procedures make their use in OBM settings a complicated endeavor (Austin, 1999).

Because of these obstacles, Austin (1999) argued that functional assessments in organizational settings should be different from the functional assessment used by ABA to identify maintaining variables for problem behavior. He described some of the work done by practitioners in OBM and proposed a model for measuring maintaining variables in organizational settings. One example is Mager and Pipe (1997), who presented a flowchart with a series of questions aimed at identifying whether there were skill deficits or issues with consequences contributing to a performance problem. This model considered some sources contributing to performance deficits, but Austin (1999) argued that more variables should be included in the assessment. Additionally, Gilbert (1978) proposed two models: vantage analysis and troubleshooting. The vantage analysis addressed issues based on disagreements between individuals within the organization, mainly looking at
information and antecedent-based variables. The troubleshooting model was meant to identify hindering variables in the environment and asked a series of questions in six areas including: (1) information; (2) instruments and equipment; (3) motivation and incentives; (4) knowledge; (5) capacity; and (6) employee motives. Rummler and Brache (1995) developed a model to analyze the organizational, process, and performer levels of an organization. Austin (1999) stated that the model was excellent for connecting organizational processes at various levels to the overall mission and strategy of the organization. However, Rummler and Brache’s (1995) model did not address the issues of motivating individual performers within the different processes of an organization. Finally, Austin (1999) discussed the ABC analysis model, as used by Daniels and Bailey (1989), which looked at antecedents and consequences surrounding behavior in both the process and performer levels of analysis. The ABC analysis served as a basepoint for Austin’s (1999) model but may have been too broad an analysis because of the potentially vast range of environmental variables surrounding performance.

Therefore, Austin (1999) organized what he considered relevant environmental variables into four areas that were demonstrated in the empirical literature to contribute to performance deficits. The four areas considered were: (a) **Antecedents**; (b) **Equipment and Processes**; (c) **Knowledge and Skills**; and (d) **Consequences**. The model also included subdomains known as “factors” within each area and “parameters” or relevant dimensions for each factor. An example would be
the “Antecedents” area that had “Priorities/Instructions” as one of its factors where the relevant parameter was whether instructions were clear, concise, and accurate. This new model was based on the models of performance analysis previously discussed as well as Gilbert’s (1982a; b) probe model, Komaki et al.’s (1986) Operant Supervisory Taxonomy Index (OSTI), and a review of OBM literature that examined deficits likely to fall into one of these areas (Austin, 1999). Each of these four main areas or domains will be discussed in further detail in the following sections.

The area of antecedents focused on factors like priorities, instructions, the mission statement, prompting, goals, and rules. These factors pertained to information provided by the organization to the individual performer and focused on clarifying, simplifying, and revising antecedent variables. This involved setting up clear instructions, prompts that were simple to understand, and goals that were regularly revised to be both challenging and attainable.

The area of equipment and processes focused on identifying maintaining/hindering variables in the functioning and physical arrangement of equipment, as well in the different levels of analysis. Austin (1999) noted that reliably analyzing equipment functioning and physical arrangement is not difficult. It can be done by monitoring the outputs produced by the equipment and through an analysis of how the physical area is arranged. When addressing the different processes involved in the organization, Rummler and Brache’s (1995) three levels of
analysis served as a guideline for identifying variables contributing to performance problems within these processes.

The knowledge and skills area focused on potential variables that contributed to either a lack of knowledge or a lack of physical skill. For this area, Austin (1999) explained that many problems encountered in work settings result from inadequate skills or lack of knowledge of the target task. Certain tasks may require more physical skills, like fine motor coordination, whereas others may require increased verbal skills.

Finally, the consequences area looked at the different variables that followed behavior. These are variables that have occurred after an instance of behavior and may contribute to the future likelihood of a particular response. Part of the questions for this area was based on Komaki et al.’s (1986) OSTI; the OSTI is used as a measure of supervisory behavior to identify aspects of effective leadership (Austin, 1999). Since much of what supervisors and managers do affects performance, Austin (1999) adapted the OTSI to include supervisory behavior as part of an organizational functional assessment.

An applied example of this model was provided by Austin (1999) where the proposed model was used to identify variables contributing to deficits in the performance of employees in a pressroom for a large newspaper. Using the model, Austin (1999) investigated potential deficits in the four main areas previously discussed. An example identified in the consequence section was that employees
were receiving limited or inconsistent feedback from their supervisors on their performance. The recommendations in this category involved daily verbal feedback from supervisors on measurable outcomes of behavior including setting up the presses, the timeliness and quality of the paper produced, and any waste produced (Austin, 1999). The use of a functional assessment allowed researchers to design interventions that target a specific performance concern and address the relevant variables contributing to that concern.

This model was expanded upon by Austin (2000). Performance analysis applies the principles of behavior analysis to aid managers and consultants in the assessment of performance concerns (Austin, 2000). In this chapter, Austin referred to the tool as the Performance Diagnostic Checklist (PDC). The PDC was designed as an informant-based tool to guide managers and consultants in identifying the need for interventions at the individual performer level. It consists of 20 questions asked around the four main areas: (1) Antecedents; (2) Equipment and processes; (3) Knowledge and skills; (4) and Consequences. Questions were designed to be answered as either YES or NO and the total number of questions answered NO are summed up at the end of the interview. The area that contained the most NOs would be considered an area for improvement and interventions were designed to specifically address the indicated concern. Austin (2000) stated that despite the lack of ‘true FAs’ in OBM, the laws of behavior are just as valid in organizational settings as in other applied settings. Therefore, the goal should be to develop and improve
methods of measuring and manipulating contingencies in these complex settings. A few studies involving the PDC will be discussed to highlight the variables identified by the tool and the interventions that were put in place to address concerns around the performance of employees.

The PDC has been used in a variety of settings including restaurants (Amigo et al., 2008; Austin et al., 2005; Rodriguez et al., 2005), coffee shops, (Pampino et al., 2004), retail stores (Pampino et al., 2003; Rohn et al., 2003), and large department stores (Eikenhout & Austin, 2005). Rohn et al. (2003) used the PDC to interview employees in a retail store that had been experiencing consistent cash shortages. The results of the functional assessment identified potential deficits in the Equipment and Processes and Consequences sections. An intervention was developed that established an accountability process that assigned each employee to a respective cash register and included a feedback package of daily verbal and graphed feedback. The results of the intervention indicated that cash shortages went from a daily average of $2.27 at baseline to $0.06 during the intervention. A reversal demonstrated that average daily cash shortages rose to $7.78 during the second baseline and decreased to $0.19 when the intervention was reinstated (Rohn et al., 2003).

Another assessment with the PDC was conducted by Pampino et al. (2004) to address deficits in the completion of maintenance tasks by employees in a coffee shop. Maintenance tasks were categorized into two groups; task group 1 consisted of
stocking duties and task group 2 consisted of cleaning duties. The results of the PDC identified deficits in the Antecedents and Consequences sections of the tool. An intervention that included task clarification and a lottery system was implemented. The average number of maintenance tasks completed increased from 44% and 32% to 86% and 67% for task groups 1 and 2, respectively (Pampino et al., 2004). These studies demonstrated that the use of pre-intervention assessments, like the PDC, can inform managers and consultants of relevant dimensions contributing to performance concerns. Interventions can then be tailored to target these concerning areas. Modifications have been made to the PDC to address performance concerns in specific settings, one of which is the human service industry (Carr et al., 2013).

According to Reid and Parsons (2000), the goal of OBM in human service settings is twofold: (1) to ensure staff has the necessary skills to perform their duties; and (2) to ensure staff can apply these skills proficiently. Carr et al. (2013) noted that despite OBM interventions having been successful in improving staff performance in human service settings, the maintenance of performance has been troublesome. Ineffective performance in human services settings can negatively impact the quality of services provided as well as the potential outcomes achieved by clients (Carr et al., 2013). The Performance Diagnostic Checklist-Human Service (PDC-HS) was designed to directly assess employee performance when providing care to clients. The original PDC was intended for use in business and industrial settings and thus there are some items not directly relevant to the needs found in human services
settings (Carr et al., 2013). In the PDC-HS, the four main areas were modified to include: (a) Training; (b) Task Clarification and Prompting; (c) Resources, Materials, and Processes; and (d) Performance Consequences, Effort, and Competition. The modified tool also included a direct observation component for gathering more objective measures of behavior. The PDC-HS was meant to aid practicing behavior analysts in understanding performance problems that do not respond to quick fixes and help develop more sensitive, targeted interventions for these problems (Carr et al., 2013). A few studies that have used the PDC-HS will be discussed to highlight their findings and their impact on the use of pre-intervention assessments in human service settings.

In the study conducted by Carr et al. (2013), the PDC-HS was used to identify variables contributing to improper cleaning of treatment rooms by staff. As part of the study, three BCBAs were interviewed at a behavior clinic that served young children diagnosed with intellectual or developmental disabilities. Deficits contributing to staff performance were identified in the Training and Performance Consequences, Effort, and Competition sections. An important component of the study was the use of ‘non indicated’ or generic interventions that were compared to interventions based on results of the PDC-HS (i.e., indicated interventions). The purpose of this inclusion was to assess the predictive validity of the tool and determine whether there is a difference in the results of an indicated intervention and a non-indicated intervention. In the study, the indicated intervention included the use
of training and graphed feedback while the non-indicated intervention included task clarification and increased availability of materials. Following the interventions, results demonstrated that there were marked performance improvements when the indicated intervention was implemented whereas the non-indicated intervention did not produce performance improvements. This difference in results when comparing indicated versus non-indicated interventions supported the predictive validity of the tool. It also suggested that interventions based on functional assessments are more effective than generic or common interventions that do not address specific deficits in environmental variables (Carr et al., 2013).

Following its introduction in 2013, the PDC-HS has been used to address different performance concerns in a variety of human service settings. A brief review conducted by Wilder et al. (2020) identified several studies that include the PDC-HS as part of a pre-intervention assessment. These studies have been conducted in autism treatment centers (Carr et al., 2013; Ditzian et al., 2015; Wilder et al., 2018), schools (Bowe & Sellers, 2018; Merritt et al., 2019), and other settings (Cymbal et al., 2020; Smith & Wilder, 2018; Wilder et al., 2019). Ditzian et al. (2015) replicated Carr et al.’s (2013) study to address a different performance concern in an autism treatment center. The performance concern was the failure to properly secure or close therapy rooms to prevent client elopement. Researchers interviewed three supervisors to assess the performance of four staff members. The results of the PDC-HS suggested that there were deficits in the Performance Consequences, Effort, and Competition
section of the tool. In this case, the intervention selected was graphed feedback and the non-indicated intervention consisted of a written prompt of the desired performance. Similar to the results produced by Carr et al. (2013), the indicated intervention was effective while the non-indicated intervention was not. Researchers noted that while other methods of functional assessment like direct observation may produce more objective results, they may not be well suited for organizational settings due to factors like rule-governance, which makes controlling variables difficult to identify (Ditzian et al., 2013).

Wilder et al. (2018) noted that previous studies using the PDC-HS had only addressed performance concerns related to the Training and Performance Consequences, Effort, and Competition sections. Their study presented two experiments; each looked at different performance concerns in an autism treatment center. The first addressed infrequent teaching of verbal operants and looked at the rate of presentation of mand, tact, and listener responding opportunities. The second addressed the irregular use of duration timers during therapy sessions. Researchers identified potential concerns in all four areas and were the first to include interventions in the areas of Resources, Materials, and Processes, and Task Clarification and Prompting. In both experiments, interventions that were indicated by the PDC-HS proved effective at increasing performance whereas interventions that were non-indicated were ineffective.
Merritt et al. (2019) used the PDC-HS in a school setting to address the tardiness of direct care staff. In this case, the PDC-HS suggested that deficits may be linked to issues in the Task Clarification and Prompting; Resources, Materials, and Processes; and Performance Consequences, Effort, and Competition sections. An intervention was implemented that consisted of task clarification, a problem-solving discussion, a token-based reward system, and graphed feedback delivered weekly. The results of the intervention demonstrated that the amount of days staff showed up on time increased for three of the four participants (Merritt et al., 2019). These studies demonstrate the importance of using pre-intervention assessments when designing interventions to increase the performance of staff members in human service settings.

Despite all the work conducted with the PDC-HS, no studies had examined the validity and reliability of the tool through 2018. Measures for reliability and validity are considered best practices for widely used tests in the fields of psychology and education (Furr & Bacharach, 2014). Validity refers to the accuracy with which the tool can perform its intended purpose. Reliability in this context refers to two things: (1) the level of agreement for scores across two or more respondents (i.e., inter-rater reliability); and (2) the level of agreement of scores produced by one person at different points in time (i.e., test-retest reliability; Wilder et al., 2019). These measures are important because the tool should be able to identify accurately and reliably what it is intended to identify. An example of measuring the validity and
reliability of informant-based tools in ABA was demonstrated in a study conducted by Iwata et al. (2013). The results obtained from the Functional Analysis Screening Tool (FAST) were compared to the results of experimental functional analysis. Since the experimental analysis is considered the gold standard for functional assessments, this comparison made sense to assess the validity of the FAST. Reliability was assessed through interrater measures that compared the results of two or more participants. Overall, the FAST score indicated the condition of the functional analysis with the highest rate of problem behavior in 63.8% of the cases reviewed (Iwata et al., 2013).

Wilder et al. (2019) stated that although there is a lack of a ‘gold standard’ for functional assessments in organizational settings, research on the validity and reliability of informant-based tools can still be conducted. As part of the study they created three video vignettes, using confederates, that depicted an interview between a consultant and direct supervisor who discuss a performance concern for an employee. Each video was designed to indicate one or more specific PDC-HS domains; the first indicated just one domain, the second video indicated two domains, and the third video indicated three domains. The study recruited twenty-one participants including seven at the associate’s level of education, seven at the bachelor’s level, and seven at the master’s level in behavior analysis or psychology; each was tested twice with a gap of two to four weeks per administration. Participants
were given the PDC-HS and asked to watch the videos while scoring the answers provided by the supervisor.

Validity was measured by whether or not participants were able to identify the correct domain(s) per video. Interrater reliability was measured by comparing participants’ scores to one another and test-retest reliability was measured by comparing two scores for one individual, conducted two to four weeks apart. The average accuracy in identifying correct domains across all participants was 97%, the average inter-rater reliability was 95.2%, and the average test-retest reliability was 97.2%. Although overall scores were high across participants, it may have been a function of how the information was provided in the vignettes. Each script presented relatively straightforward answers, which is not a likely occurrence in an organizational setting. Wilder et al. (2019) described the study as an important first step in measuring the validity and reliability of the PDC-HS; had the outcome in these contrived conditions been poor, the effort to measure validity and reliability in more complex conditions would not have been necessary.

Cymbal et al. (2020) sought to address this issue of external validity by basing the video vignettes on actual interviews that used the PDC-HS to address real-world performance concerns. Their study replicated the work of Wilder et al. (2019) with a modification to the video vignettes. Similar to Wilder et al. (2019), Cymbal et al. (2020) assessed validity and reliability. Participants were instructed to watch the vignettes and score them accordingly; they were also asked to complete the PDC-HS
again two to four weeks after the first session. Results demonstrated a slight reduction in accuracy where the average across participants was 86%; interrater reliability and test-retest reliability scores were 79% and 83%, respectively. It was expected that a less contrived script that more closely resembled real-world interviews would affect the accuracy and reliability of the tool, and this hypothesis was confirmed. However, the results were still promising and suggest that the PDC-HS can accurately measure what it is intended to measure (Cymbal et al. 2020). A limitation that was noted in the studies by Wilder et al. (2019) and Cymbal et al. (2020) was that participants were not required to select an intervention following the completion of the PDC-HS. This is a necessary step in the functional assessment process where an intervention is selected based on the results of the assessment. Without this information, it is unclear whether participants who were able to accurately complete the tool, based on the information in the vignettes, would then be able to select an appropriate intervention.

Recently, a modified version of the PDC-HS was released as the PDC-HS 1.1. The PDC-HS 1.1 was designed to improve upon certain aspects of the previous version. Some of the modifications include an extension of the instructions for using the tool which provides additional information to the interviewer on how to conduct the informant-based assessment. The revised version also includes expansions to the Training and Resources, Materials, and Processes sections to include additional information about training types as well as an emphasis on conducting direct
observations to identify equipment use and arrangement. Additionally, some questions were reworded so that an answer of NO demonstrates an issue in the relevant area. For example, one question on the original PDC-HS asked “Do other tasks appear to take precedence over the target task? If yes, indicate these tasks below.” Since the scoring for the PDC-HS is designed to identify deficits based on the number of questions answered NO per relevant area; an answer of YES for this question would not be indicated as a concern despite there being competing contingencies interfering with the effective performance of the targeted task. The PDC-HS 1.1 rephrases this question to ask, “Does the task generally take precedence over potentially competing tasks?” Answering NO to this question would indicate that there are competing tasks that interfere with the completion of the targeted task. This provides a clearer method of identifying a performance concern in the Performance Consequences, Effort, and Competition section of the tool.

The purpose of the present study was to replicate the work performed by Wilder et al. (2019) and Cymbal et al. (2020) to assess the PDC-HS 1.1. Three video vignettes, each depicting an interview conducted using the PDC-HS 1.1, were administered. Similar to Cymbal et al. (2020), the vignettes were based on real-world interviews to strengthen external validity. The current study also included an intervention selection component following completion of the PDC-HS 1.1; this has not been done previously. Participants were asked to score the interviews using the PDC-HS 1.1 and select an intervention based on the indicated section(s).
Chapter 2
Method

Participants, Setting, and Materials

Twenty-one individuals participated in the study. Consistent with Wilder et al. (2019) and Cymbal et al. (2020), we recruited participants across experience levels to measure variations in PDC-HS 1.1 scoring accuracy. We included seven participants with a Master’s degree in psychology or a related field, seven participants with a Bachelor’s degree in psychology or related field, and seven participants who were undergraduate students with a high school diploma or equivalent. The Master’s level group included 6 females and 1 male participant with an average age of 31 years ($SD = 7.81$) and an average of 9.21 years ($SD = 5.37$) of experience in Behavior Analysis; 6 of 7 were Board Certified Behavior Analysts® (BCBAs®). The Bachelor’s level group included 6 females and 1 male participant with an average age of 33.57 years ($SD = 5.47$) and an average of 4.21 years ($SD = 3.01$) of experience in Behavior Analysis; 1 of 7 was a Board Certified Behavior Assistant Analyst® (BCaBA®). The undergraduate group included 5 females and 2 male participants with an average age of 21.14 years ($SD = 1.68$) and an average of 0.42 years ($SD = 0.78$) of experience in psychology or Behavior Analysis; only 2 of the 7 participants in the undergraduate group had received formal training in behavior
analysis. Data was not gathered on whether participants had received a course in OBM prior to the start of the study.

Participants were recruited from private clinics that provide services to children with autism and from graduate and undergraduate cohorts of universities in the southeastern and midwestern United States. Participants were recruited based on their familiarity with psychology and behavior analysis but none had exposure to the PDC-HS 1.1 before the start of the study. Sessions were conducted online via encrypted video communications software (ZOOM®). Materials included a laptop computer with speakers to view videos and digital copies of the PDC-HS 1.1 (See Appendix).

**Video Vignettes**

We created three different vignettes; each was approximately 8-14 minutes long and indicated a performance concern across one or more PDC-HS 1.1 areas. As in Wilder et al. (2019) and Cymbal et al. (2020), each vignette was structured as an interview between a consultant using the PDC-HS 1.1 and a supervisor who answered questions based on the performance concern of an employee. Each vignette also included a 1–2-minute observation of the particular employee demonstrating the performance deficit; this was meant to allow scoring of questions that required an observation component to be verified. Similar to Cymbal et al. (2020) the video scripts were based on six different real-world interviews in which the PDC-HS 1.1
was used with a supervisor. Each script was arranged to indicate a different area(s) of the PDC-HS 1.1 as contributing to a performance concern. The roles of consultant, supervisor, and employee were played by graduate student confederates. To assess validity across increasingly complex performance concerns the vignettes indicated an increasing number of PDC-HS 1.1 areas; the first vignette indicated one area, the second vignette indicated two areas, and the third indicated three areas.

An intervention selection component was included based on the recommendations offered in Wilder et al. (2019) and Cymbal et al. (2020). The interventions for each indicated domain were selected based on the suggestions provided in the Intervention Planning section on page 7 of the PDC-HS 1.1. Each of these interventions is accompanied by one or more literature citations that illustrate the implementation of the intervention in an applied setting. The first vignette indicated the Resources, Materials, and Processes domain and the selected intervention was to improve access to, redesign, and reorganize task materials. The second vignette indicated the Training and Task Clarification and Prompting domains; the related interventions were behavioral skills training and task clarification and checklists. Lastly, the third vignette indicated: 1) Performance Consequences, Effort, and Competition; 2) Training; and 3) Task Clarification and Prompting. The interventions selected for the third vignette were: 1) increased supervisor presence; 2) improve access to, redesign, and reorganize task materials; and 3) task clarification and checklists.
Data Collection and Dependent Variables

As a measure of the validity of the PDC-HS, we assessed whether participants were able to correctly identify the PDC-HS 1.1 domain indicated as a concern in each vignette (i.e., the area with the most “no” answers) and select an appropriate intervention. We calculated the number of questions that were answered correctly, in each vignette, according to the information provided on employee performance in the video. We also calculated the average percentage of correct answers across each vignette and each PDC-HS 1.1 domain. To calculate the validity of the intervention selection component, we calculated the average percent correct selection across and within each group (i.e., master’s, bachelor’s, and pre-bachelor’s).

To calculate test-retest reliability, we compared two completed PDC-HS 1.1 assessments for the same participant at the start and end of the study, which was between two and four weeks apart. We calculated percent agreement by comparing each participant’s responses on each item of the PDC-HS during their first session to the answers provided during their second session. We then calculated the total agreement per participant by dividing the number of agreements by the number of agreements plus disagreements, multiplying this number by 100 and converting it to a percentage. Finally, we calculated the percent average test-retest agreement across all the participants.
To calculate interrater agreement, participants were randomly paired within each group and responses for each item on the PDC-HS were compared. For each pair, we divided the total number of agreements by the total number of agreements plus disagreements, multiplied this by 100 and converted it to a percentage. Comparisons were made using an analysis of variance.

Procedure

Participants signed the informed consent form prior to their online session, each participant provided their own laptop computer on which they worked, these were the only materials required. The researcher followed a task script adapted from Cymbal et al. (2020). In the script, the researcher provided a verbal description of the PDC-HS 1.1 and then allotted five minutes for the participant to familiarize themself with the tool’s questions and observation requirements. The experimenter maintained full control of the videos and did not pause or playback any portions; each video was played once per session in its entirety. When the participant finished the first video, the researcher allotted an additional three minutes for the intervention selection component where the participant selected an intervention based on the results of the assessment. Once this time expired, the researcher restated the instructions without the description of the PDC-HS 1.1 or the familiarization period and played the second video. The same process was repeated for the third video and
once the participant finished the third video, the session ended. The videos were presented in a random sequence for each participant. Researchers did not answer any clarifying questions regarding the content of the video. Approximately two to four weeks after the first session, participants were provided the same instructions, materials, and vignettes to score a second PDC-HS 1.1 for each video. Similar to the first session, participants were not able to stop, replay, or fast-forward the videos. After completing the second session, participants were compensated between 10 and 20 U.S. dollars for their participation.
Chapter 3
Results

Validity

Most of the participants were able to accurately identify the domain(s) of the PDC-HS 1.1 responsible for the performance concern in each vignette ($M = 90.36\%, SD = 5.6$). Overall accuracy in the intervention-selection component was $79.76\%$ ($SD = 13.58$). In the first video, which had Resources, Materials, and Processes as an indicated domain, participants accurately answered an average of $91.64\%$ ($SD = 4.61$) of the questions. The accuracy of selection for the intervention component in the first video was $95.24\%$ ($SD = 15.04$). In the second video, which indicated both the Training and Task Clarification and Prompting domains, participants accurately answered $91.56\%$ ($SD = 9.75$) of the questions. The accuracy of selection for the intervention component in the second video was $84.52\%$ ($SD = 24.34$). In the third video, in which three domains were indicated (i.e., Training, Task Clarification and Prompting, and Performance Consequences, Effort, and Competition), participants accurately answered $87.77\%$ ($SD = 6.64$) of the questions. The accuracy of selection for the intervention component in the third video was $71.02\%$ ($SD = 18.46$). Overall, participants scored better in video one than they did in video two and scored better
in video two than in video three; this trend is repeated for scores of the intervention-selection component (see Table 4).

The Master’s group had an overall accuracy score of 92.51% ($SD = 4.39$), the Bachelor’s group had a score of 90.15% ($SD = 5.18$), and the Pre-bachelor’s group had a score of 88.31% ($SD = 6.9$). For the Training domain, participants correctly answered an average of 90.63% ($SD = 6.96$) of questions. In the Task Clarification and Prompting domain, participants answered 93.81% ($SD = 6.85$) of questions correctly. In the Resources, Materials, and Processes domain, participants correctly answered 85.26% ($SD = 8.54$) of the questions. Participants answered 93.17% ($SD = 6.54$) of questions correctly in the Performance Consequences, Effort, and Competition domain (see Table 1).

Table 1 - PDC-HS 1.1 Scoring Summary

<table>
<thead>
<tr>
<th>Summary of PDC-HS 1.1 Accuracy</th>
<th>Vignette 1</th>
<th>Vignette 2</th>
<th>Vignette 3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Overall</td>
<td>91.64</td>
<td>4.61</td>
<td>91.56</td>
<td>9.75</td>
</tr>
<tr>
<td>By Participant Experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s</td>
<td>92.79</td>
<td>2.76</td>
<td>96.1</td>
<td>7.51</td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>90.26</td>
<td>5.97</td>
<td>89.93</td>
<td>6.92</td>
</tr>
<tr>
<td>Pre-Bachelor’s</td>
<td>91.88</td>
<td>4.88</td>
<td>88.63</td>
<td>13.25</td>
</tr>
<tr>
<td>By Domain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task Clarification and Prompting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resources, Materials, and Processes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance Consequences, Effort, and Competition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

To examine the effects of education/experience level (i.e., Pre-Bachelor’s, Bachelor’s degree, and Master’s degree) and video complexity (i.e., one, two, and three indicated problematic domains) on video scoring accuracy, we conducted a 3
(experience level) x 3 (video complexity) factorial ANOVA. There were no significant differences found in scores between experience groups, $F(2, 54) = 1.770, p = .180$, or across video complexity, $F(2, 54) = 1.957, p = .151$. 

To examine the effects of education/experience level and video complexity on intervention-selection accuracy, we conducted a 3 (experience level) x 3 (video complexity) factorial ANOVA. There was a significant main effect for video complexity, $F(2, 54) = 8.04, p < .001$, partial $\eta^2 = .23$. Post-hoc tests using Tukey HSD suggested that intervention-selection accuracy was significantly lower when participants had to select three interventions ($M = 71.02\%$) than when selecting just one intervention ($M = 95.24\%$). There were no significant differences found in scores between experience groups, $F(2, 54) = 2.66, p = .079$.

**Test-retest reliability**

The overall mean test-retest reliability score for the 21 participants was 87.8\% ($SD = 8.39$) (see Table 2). The Master’s level group had the highest test-retest reliability score with an average of 90.69\% ($SD = 6.57$). The Bachelor’s level group showed an average test-retest score of 86.79\% ($SD = 7.61$). The undergraduate group had the lowest score at 85.93\% ($SD = 10.88$) test-retest reliability. For the test-retest reliability, video one (i.e., one domain indicated) had the highest score at 90.9\% ($SD = 6.43$). The second video (i.e., two domains indicated), had a test-retest reliability
score of 87.66% (SD = 14.09). Finally, the third video (i.e., three domains indicated) had a test-retest reliability score of 84.84% (SD = 10.1).

Table 2 - Summary of PDC-HS 1.1 Test-retest Reliability

<table>
<thead>
<tr>
<th>Variable</th>
<th>Vignette 1 M</th>
<th>Vignette 1 SD</th>
<th>Vignette 2 M</th>
<th>Vignette 2 SD</th>
<th>Vignette 3 M</th>
<th>Vignette 3 SD</th>
<th>Overall M</th>
<th>Overall SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overal</td>
<td>90.9</td>
<td>6.43</td>
<td>87.66</td>
<td>14.09</td>
<td>84.84</td>
<td>10.1</td>
<td>87.8</td>
<td>8.39</td>
</tr>
</tbody>
</table>

By Participant Experience

<table>
<thead>
<tr>
<th>Experience</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master's</td>
<td>93.5</td>
<td>3.58</td>
</tr>
<tr>
<td>Bachelor's</td>
<td>90.26</td>
<td>6.11</td>
</tr>
<tr>
<td>Pre-Bachelor's</td>
<td>88.96</td>
<td>8.65</td>
</tr>
</tbody>
</table>

Table 3 - Summary of PDC-HS 1.1 Interrater Reliability

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>85.92</td>
<td>6.58</td>
</tr>
</tbody>
</table>

By Participant Experience

<table>
<thead>
<tr>
<th>Experience</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Master's</td>
<td>89.01</td>
<td>6.62</td>
</tr>
<tr>
<td>Bachelor's</td>
<td>86.17</td>
<td>7.86</td>
</tr>
<tr>
<td>Pre-Bachelor's</td>
<td>82.57</td>
<td>5.06</td>
</tr>
</tbody>
</table>

By Vignette

<table>
<thead>
<tr>
<th>Vignette</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vignette 1</td>
<td>89.39</td>
<td>4.97</td>
</tr>
<tr>
<td>Vignette 2</td>
<td>86.74</td>
<td>9.88</td>
</tr>
<tr>
<td>Vignette 3</td>
<td>81.63</td>
<td>8.36</td>
</tr>
</tbody>
</table>

Interrater reliability

The overall interrater reliability score, calculated from random participant pairings within each experience group, was 85.92% (SD = 6.58) (see Table 3). Similarly to Cymbal et al. (2020), interrater agreement scores varied across
the three groups, with the Pre-Bachelor’s group ($M = 82.57\%, SD = 6.58$) being lower than both the Master’s group ($M = 89.01\%, SD = 6.62$) and the Bachelor’s group ($M = 86.17\%, SD = 7.86$). Interrater reliability also varied between each video; agreement for video three ($M = 81.63\%, SD = 8.36$) was lower than both video one ($M = 89.39\%, SD = 4.97$) and video two ($M = 86.74\%, SD = 9.88$).

**Table 4 - Summary of Intervention-Selection Accuracy**

<table>
<thead>
<tr>
<th>Variable</th>
<th>$M$</th>
<th>$SD$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall</strong></td>
<td>79.76</td>
<td>13.58</td>
</tr>
<tr>
<td><strong>By Participant Experience</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s</td>
<td>86.9</td>
<td>8.13</td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>80.95</td>
<td>11.5</td>
</tr>
<tr>
<td>Pre-Bachelor’s</td>
<td>71.43</td>
<td>16.57</td>
</tr>
<tr>
<td><strong>By Vignette</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vignette 1 (one intervention)</td>
<td>95.24</td>
<td>15.04</td>
</tr>
<tr>
<td>Vignette 2 (two interventions)</td>
<td>84.52</td>
<td>24.34</td>
</tr>
<tr>
<td>Vignette 3 (three interventions)</td>
<td>71.02</td>
<td>18.46</td>
</tr>
</tbody>
</table>
We assessed measures of validity and reliability of the PDC-HS 1.1, as a replication and extension of Wilder et al. (2019) and Cymbal et al. (2020). To assess validity, participants watched three video vignettes that simulated an interview, regarding a performance concern, while completing the PDC-HS 1.1. Each vignette indicated an increasing number of domains across video 1 (i.e., one domain indicated), video 2 (i.e., two domains indicated), and video 3 (i.e., three domains indicated).

The first video indicated the Resources, Materials, and Processes domain and the intervention prioritized for this domain was to improve access to, redesign, and reorganize task materials. The second video indicated the Training and Task Clarification and Prompting domains; behavioral skills training and task clarification and checklists were selected as the respective interventions for video 2. The third vignette had: 1) Performance Consequences, Effort, and Competition, 2) Training, and 3) Task Clarification and Prompting) as indicated domains. The relevant interventions selected for video 3 were: 1) increased supervisor presence; 2) improve access to, redesign, and reorganize task materials; and 3) task clarification and checklists. Most participants correctly identified the domain(s) responsible for the performance deficit depicted in the videos (90.36%) and selected...
a relevant intervention (79.76%). We found no statistically significant difference in scoring accuracy across the three videos but we did find a significant difference in intervention-selection accuracy where participants scored significantly lower on video 3 than on video 1.

The accuracy of participant groups across domains and video vignettes varied somewhat, with the Master’s group having the highest overall score. The Master’s group had the highest scores for both the first and second videos and the second-highest score for the third video. The Pre-bachelor’s groups had the lowest overall accuracy score and the lowest scores across all three video vignettes, although none of the differences were statistically significant. These results are similar to those produced by Cymbal et al. (2020) and suggest that experience in behavior analysis only slightly impacts the accuracy of scores produced when completing the PDC-HS 1.1. The first video had the highest accuracy score of all three videos and the third video had the lowest accuracy score. This was expected because the vignettes were designed to be increasingly complex in terms of indicated domains. The domain with the lowest accuracy score was Resources, Materials, and Processes which is similar to the results produced by Cymbal et al. (2020).

Values for test-retest reliability and interrater reliability were overall high. Similar to the accuracy scores, the Master’s group had the highest overall test-retest reliability and interrater reliability scores. The intervention-selection scores also varied across videos and participant groups. The first video produced the highest
intervention-selection accuracy score and the third video produced the lowest score. These results are similar to those produced in the scoring portion and are likely due to similar reasons in that the number of indicated domains, and their relevant interventions, increased across videos. Since the third video had three indicated domains, this may have increased the difficulty of selecting a relevant intervention for each domain. The Master’s group had the highest overall accuracy in the intervention selection component. The overall score of the intervention-selection component was lower than the overall accuracy score.

The results of this study suggest that participants, regardless of experience level, can accurately and reliably implement the PDC-HS 1.1 to identify a performance concern and select a relevant intervention. Although participants in the Pre-Bachelor’s group scored the lowest overall, their accuracy scores for the scoring portion were between 80-90% and the average intervention-selection score was above 70%. The results also provide further support for the use of informant-based methods for conducting functional assessments to identify employee performance concerns.

Functional assessments can involve informant-based records reviews and interviews, direct observations of behavior in its natural setting, or experimental manipulations of environmental variables. An experimental functional analysis is the only method that can identify causal variables controlling behavior. When an experimental analysis is impractical, informant-based and descriptive assessment
methods are used to approximate the function of a behavior. The PDC-HS 1.1 combines informant-based assessment with direct observation to measure and record employee performance as it occurs. These observation components help supplement the tool and increase its utility in identifying obstacles in the work setting contributing to a particular performance concern.

In real-world scenarios, participants will be able to review the tool and practice before implementation. They will also be able to conduct the observations themselves which may improve the effectiveness of the tool. A practitioner with training in behavior analysis might be better able to ask appropriate follow-up questions and prioritize relevant interventions. Based on the methods described in Wilder et al. (2019) and Cymbal et al. (2020), this study provides support for the utility of the PDC-HS 1.1 in identifying environmental variables contributing to performance concerns and selecting a relevant intervention to prioritize.

Anecdotally, one of the more common sections where participants scored incorrectly was in the Resources, Materials, and Processes domain where question #2 asks “If materials are required for task completion, are they readily available?” The supervisor answered, “Yes, materials are required” the consultant then asked a follow-up question “So, are they readily available”, to which the supervisor answered “No.” This was a common question that was answered as “Yes” when the correct answer was scored as “No”
This study had some limitations. Several of the limitations are similar to those of Cymbal et al. (2020). The first limitation is the lack of a ‘gold standard’ with which to compare the outcomes of informant tools in OBM. The empirical literature in applied behavior analysis has demonstrated repeatedly, through replications and extensions, that an experimental functional analysis of the type conducted in clinical behavior analysis, is the ‘gold standard’ for identifying operant contingencies controlling behavior. However, the field of OBM lacks such a standard with which to compare assessments that do not perform experimental manipulations. This scarcity is likely due to several reasons including the fact that interventions in OBM have had relatively high success without using experimental analysis, the contingencies controlling behavior in a work setting may not be readily apparent, and a formal functional analysis may be too costly and impractical in a work setting. Thus, any assessment of the validity of informant-based performance analysis tools will require contriving scenarios depicting a particular outcome with which to compare the results of assessments completed by participants. Informant-based tools are a common functional assessment method in OBM and the literature available for the PDC-HS demonstrates the utility of the tool in identifying problematic areas contributing to performance deficits and selecting relevant interventions to improve performance.

The second limitation is the size and scope of the participant pool for this study. Since there were only seven participants included per group, the
generalizability of the results obtained is limited. A third limitation is that the participants were not randomly selected to participate and were not randomly assigned. Instead, participants were assigned to three different groups based on their level of experience. The purpose of doing so was to compare the PDC-HS 1.1 scores of participants across different experience levels. Future research on the validity and reliability of the PDC-HS 1.1 should consider larger sample sizes, random selection, and random assignment.

A fourth limitation is that participants were watching an assessment being conducted by a consultant and following along based on the answers provided in the videos. That is, the participants did not perform the assessment themselves but watched a consultant interview a supervisor. The PDC-HS 1.1 is meant to be completed directly by the consultant as they interview the supervisor or performer. Future research should consider using a live confederate trained with responses for PDC-HS 1.1 questions so that participants can conduct the interview themselves. Another suggestion for future research could be a scripted scenario that allows participants to ask or select pre-determined follow-up questions that provide additional information related to a particular performance concern and the answers given on the PDC-HS 1.1.

Another limitation was test sensitization; participants completed identical assessments for identical vignettes. Although the vignettes were presented in a random sequence each session, any improvements in participants’ scores during the
second session were likely due to having had previous exposure to the PDC-HS 1.1 and the video vignettes. This might also skew test-retest reliability scores for participants who had marked improvements from one session to the next. For example, participant #4 went from 66.67% overall accuracy in the first session to 83.33 in the second session; their overall test-retest score was the lowest at 65.15%.

Finally, one limitation unique to this study may lie in the intervention selection component itself and the lack of an established standard for prioritizing an intervention within an indicated domain. For this study, an intervention was chosen for each of the indicated domains across the three vignettes. The first vignette, which indicated one domain as a priority (i.e., Resources, Materials, and Processes), had only one intervention selected (i.e., improve access to, redesign, and reorganize task materials). The second vignette indicated two domains and had behavioral skills training and task clarification and checklists as the two selected interventions. Lastly, the third vignette indicated three domains and had three interventions selected: 1) increased supervisor presence; 2) improve access to, redesign, and reorganize task materials; and 3) task clarification and checklists).

Selection of the interventions was based on the recommendations in the Intervention Planning section of the PDC-HS 1.1 which provides sample interventions based on particular questions that were answered as “No” within any of the domains. An example of this can be seen in the Performance Consequences, Effort, and Competition domain where an answer of “No” to question #1 (i.e., Is the
employee ever directly monitored by a supervisor when the target task is to be performed?) will indicate increased supervisor presence as a potential intervention. Although it makes sense to increase supervisor presence if the employee is not receiving supervision, there is no objective standard to determine whether increasing supervisor presence is appropriate, how to do so properly, or whether the intervention should be considered a priority at all. Generally, there is no objective standard for which intervention to prioritize in an indicated domain. The limited experience in behavior analysis of participants in the Pre-Bachelor’s group (i.e., only 2 of 7 had formal training) may have affected their accuracy scores in the intervention-selection component. Since the recommended interventions in the “Intervention Planning” section are based on the empirical literature in applied behavior analysis, individuals without formal training in behavior analysis may lack the expertise necessary to identify a relevant intervention based on the environmental obstacles indicated in a particular domain of the PDC-HS 1.1.

Vance et al. (2022) offered procedural refinements to the PDC-HS and identified aspects of the intervention-selection process that could inhibit the objectivity and utility of the tool. The researchers identified the lack of a clear cutoff threshold for identifying a priority domain and the lack of a decision-making model to prioritize a single domain as the focus of the intervention. These aspects may have limited the objectivity of the intervention-selection component for our study and impacted the validity of the scores produced. Vance et al. (2022) proposed a cutoff
threshold for identifying priority domains and provided several decision-making models for prioritizing a domain for intervention when multiple domains are indicated. For the cutoff threshold, the researchers used a norm-based standard-setting method by reviewing existing articles where a PDC-HS intervention was implemented. Researchers identified the point at which the PDC-HS administrators determined a particular domain warranted intervention. The cutoff threshold was calculated based on the percentage of questions answered “No” (i.e., negativity score) in a particular domain for which an intervention was selected. The cutoff threshold was set at a 50% negativity score which means that at least 50% of the questions in a particular domain had to score as a “No” for the domain to be considered a priority.

Although each of the indicated domains in the video vignettes met this cutoff threshold, it was not formally used to select the priority domains and was not provided to participants when making their selections. The PDC-HS 1.1 provides instructions to prioritize the most problematic areas or to give priority to areas in which multiple items are endorsed. This however does not provide a clear indication of which domain to prioritize when multiple items are endorsed in more than one domain. The cutoff threshold proposed by Vance et al. (2022) may serve as an initial step for establishing a standard for prioritizing PDC-HS 1.1 domains. Future research with the PDC-HS 1.1 should consider the use of this cutoff threshold when identifying domains to prioritize for intervention.
The purpose of the decision-making model proposed by Vance et al. (2022) is to help practitioners identify a single domain as the focus of intervention when two or more domains are indicated. The decision-making model is based on previous applications of the PDC-HS, not the PDC-HS 1.1, and although there are similarities, some aspects of the PDC-HS 1.1 have been modified so that the decision-making model proposed for the PDC-HS does not readily apply to the modified tool. For our study, participants were asked to select an intervention for each of the indicated domains per vignette (i.e., one, two, and three domains) but were not asked to prioritize an intervention within a vignette when multiple domains were indicated as a concern (e.g., vignette 2 and vignette 3). Future research with the PDC-HS 1.1 should consider the update and use of the decision-making model when investigating the intervention-selection component of the assessment. Participants may be asked to prioritize an intervention, while using the decision-making model, in situations where two or more domains are indicated.

To summarize, this study provides support for the validity and reliability of the PDC-HS 1.1 when measured in the manner described herein. The results suggest that the PDC-HS 1.1 may be used more effectively by individuals who are trained in behavior analysis but the difference in scores across experience levels was not markedly high. The scores obtained in the intervention selection component suggest that, despite a lack of objective standards for intervention selection, participants were generally able to select a relevant intervention for each of the indicated domains.
Future research should focus on identifying and teaching the skills necessary to effectively implement the assessment.


https://doi.org/10.1300/j075v25n03_02

https://doi.org/10.1300/j075v22n01_03


Performance Diagnostic Checklist - Human Services (1.1)

Employee Name: 
Supervisor Name: 
Assessor Name: 
Describe Performance Concern:

INSTRUCTIONS
*For a comprehensive user guide regarding this assessment, see (INSERT ARTICLE LINK HERE)

1. Conduct as an interview with the employee’s direct supervisor. Do not simply hand it to the supervisor and ask them to complete by themselves, unless they are well trained in completing the PDC-HS (1.1).
2. Answer the questions below about the employee’s specific performance concern (not the employee in general). The problem should be operationally defined as either a behavioral excess or deficit.
3. Complete for only one performance concern at a time. Conduct a new PDC-HS (1.1) for a different performance concern if needed.
4. Answer the questions in the order in which they appear.
5. Items with an asterisk (*) should be answered only after the information is verified through direct observation or interview with the employee.

Tips for conducting direct observations:
• Attempt to conduct observations during typical conditions (e.g., the employee is not sick or unduly affected by things like low staffing or unfamiliar clients)
• Conduct observations during times when the employee is typically expected to be performing the task
• Be aware of employee reactivity; employees may behave differently when they know they are being observed
• Conduct only as many observations as necessary to obtain reliable information

6. If the direct supervisor is unsure about what is being asked in a question, provide clarification or examples.
7. Answering NO vs. N/A.
   • Answer NO when the information required by the question is not an obvious/immediate “Yes”
   • Answer N/A when the information required by the question does not apply or is irrelevant to the performance concern being assessed

TRAINING

PDC-HS (1.1) – [AUTHORS TO GO HERE]
<table>
<thead>
<tr>
<th></th>
<th>Yes ☐  No ☐</th>
<th>Has the employee received formal training on this task? If YES, when did the employee receive the training?</th>
</tr>
</thead>
</table>

If the supervisor answered NO to Question 1, proceed to Question 2. If the supervisor answered YES to Question 1, check all applicable training methods below:

1a. Instructions Yes ☐  No ☐  
*The employee received step-by-step instructions on how to perform the task.*
- Vocal instructions
- Written instructions
- Both vocal and written instructions

1b. Demonstration/Modeling Yes ☐  No ☐  
*The employee was shown how to perform the task.*
- Live modeling
- Video modeling
- Both live and video modeling

1c. Rehearsal/Practice Yes ☐  No ☐  
*The employee had opportunities to practice the task correctly during training.*
- Practiced with trainer
- Practiced alone
- Practiced with coworkers

1d. Feedback Yes ☐  No ☐  
*The employee received feedback about performance during training.*
- Vocal feedback
  - Positive feedback for steps performed correctly
  - Corrective feedback for steps performed incorrectly
- Written feedback
  - Positive feedback for steps performed correctly
  - Corrective feedback for steps performed incorrectly

2 | Yes ☐  No ☐ | Is there evidence that the trainer currently responsible for training staff can accurately perform the task being trained? |

3* | Yes ☐  No ☐ | Can the employee accurately describe the target task and when it should be performed? |

4 | Yes ☐  No ☐ | Is there evidence that the employee has accurately completed the task in the past? |

5* | Yes ☐  No ☐  N/A ☐ | If the task needs to be completed quickly, can the employee perform it at the appropriate speed? (i.e., if you asked the employee to perform the task, can they do so quickly and finish it before a certain amount of time has elapsed? For example, if a therapist had five minutes to pick up all the toys on the floor before the start of the next session, could they do so within the allotted time?) |

**TASK CLARIFICATION & PROMPTING**
<table>
<thead>
<tr>
<th></th>
<th>Yes □ No □</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Has the employee been informed that they are expected to perform the task?</td>
</tr>
<tr>
<td>2*</td>
<td></td>
<td>Can the employee state the purpose of the task?</td>
</tr>
<tr>
<td>3*</td>
<td></td>
<td>Is a job aid (e.g., checklist, data sheet, step-by-step instructions, pictures, prompts) for completing the task visibly located (if appropriate) or easily accessible in the task area? (e.g., visibly posted, located with the instructional or data collection materials).</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>Is the employee ever verbally, textually, or electronically reminded to complete the task? (e.g., reminder emails, text messages, phone alerts, group messaging apps, calendar invites, verbal reminders/prompts).&lt;br&gt; If YES, how frequently is the employee reminded to complete the task?&lt;br&gt; □ Each time the task needs to be performed&lt;br&gt; □ Hourly&lt;br&gt; □ Daily&lt;br&gt; □ Weekly&lt;br&gt; □ Monthly&lt;br&gt; □ Other</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>Is the task being performed in an environment well-suited for task completion? (i.e., has the employee told you that they cannot perform the task because there is some aspect of the work environment prohibiting them from doing so. For example, if the task requires a quiet environment has the employee ever mentioned that there is too much noise, or if the task requires there to be a lot of physical space, has the employee ever stated that their work area is too crowded?).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Yes □ No □</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>Are there sufficient numbers of trained staff available in the organization to complete the task?</td>
</tr>
<tr>
<td>2*</td>
<td></td>
<td>If materials (e.g., teaching stimuli, preferred items) are required for task completion, are they readily available (e.g., easy to find, nearby)?&lt;br&gt; <em>If no materials are required (i.e., N/A) proceed to Question 6</em>&lt;br&gt; List materials below and indicate their availability:</td>
</tr>
<tr>
<td></td>
<td>N/A □</td>
<td>Item 1&lt;br&gt; Item 2&lt;br&gt; Item 3&lt;br&gt; Item 4</td>
</tr>
<tr>
<td>3*</td>
<td></td>
<td>Are there times during the day when the materials required for task completion are not readily available?&lt;br&gt; List times at which materials are unavailable:</td>
</tr>
<tr>
<td></td>
<td>N/A □</td>
<td>Item 1&lt;br&gt; Item 2&lt;br&gt; Item 3&lt;br&gt; Item 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>Yes ☐ No ☐</td>
<td>Has the employee been informed that they are expected to perform the task?</td>
</tr>
<tr>
<td>2*</td>
<td>Yes ☐ No ☐</td>
<td>Can the employee state the purpose of the task?</td>
</tr>
<tr>
<td>3*</td>
<td>Yes ☐ No ☐</td>
<td>Is a job aid (e.g., checklist, data sheet, step-by-step instructions, pictures, prompts) for completing the task visibly located (if appropriate) or easily accessible in the task area? (e.g., visibly posted, located with the instructional or data collection materials).</td>
</tr>
</tbody>
</table>
| 4 | Yes ☐ No ☐ | Is the employee ever verbally, textually, or electronically reminded to complete the task? (e.g., reminder emails, text messages, phone alerts, group messaging apps, calendar invites, verbal reminders/prompts).  
If YES, how frequently is the employee reminded to complete the task?  
☐ Each time the task needs to be performed  
☐ Hourly  
☐ Daily  
☐ Weekly  
☐ Monthly  
☐ Other |
| 5 | Yes ☐ No ☐ | Is the task being performed in an environment well-suited for task completion?  
(i.e., has the employee told you that they cannot perform the task because there is some aspect of the work environment prohibiting them from doing so. For example, if the task requires a quiet environment has the employee ever mentioned that there is too much noise, or if the task requires there to be a lot of physical space, has the employee ever stated that their work area is too crowded?). |

## RESOURCES, MATERIALS, & PROCESSES

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes ☐ No ☐</td>
<td>Are there sufficient numbers of trained staff available in the organization to complete the task?</td>
</tr>
</tbody>
</table>
| 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 (i.e., N/A) proceed to Question 6  
List materials below and indicate their availability:  
Item 1  
Item 2  
Item 3  
Item 4 |
| 3* | Yes ☐ No ☐  
N/A ☐ | Are there times during the day when the materials required for task completion are not readily available?  
List times at which materials are unavailable:  
Item 1  
Item 2  
Item 3  
Item 4 |
<table>
<thead>
<tr>
<th></th>
<th>Yes □ No □</th>
<th>N/A □</th>
</tr>
</thead>
</table>
| 4* | Are the materials necessary to complete the task well designed for their intended purpose?  
(i.e., has the employee ever told you that the materials provided to complete the task are not appropriate or useful? For example, employees may be using materials that are old and worn down and no longer suitable for efficient completion of the task, or the data sheet may be overly complicated and takes too long to complete). |
| 5* | Are the materials necessary to complete the task well organized for their intended purpose?  
(i.e., has the employee ever told you that they cannot easily and readily find and/or locate the materials needed to complete the task, or that the materials needed to complete the task are not organized in a way that assist task completion?). |
| 6 | Can the task be completed without first completing other tasks? If NO, indicate below the tasks that must be completed first:  
(i.e., has the employee ever told you that they cannot complete this task because they have to first complete other tasks? For example, the employee cannot graph the data regarding client behavior prior to calculating the daily percentages).  
Task 1  
Task 2  
Task 3  
Task 4 |
| 7 | If you answered NO for Question 6, are other employees responsible for completing any of the earlier tasks in the process? If so, indicate the employee(s) below:  
(i.e., the employee cannot complete this task without another employee completing a prior task. For example, the employee is unable to file paperwork on time because another employee has to complete proofreading and editing).  
Task 1  
Task 2  
Task 3  
Task 4 |

**PERFORMANCE CONSEQUENCES, EFFORT, & COMPETITION**

<table>
<thead>
<tr>
<th></th>
<th>Yes □ No □</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Is the employee ever directly monitored by a supervisor when the target task is to be performed? If YES, indicate the frequency of monitoring:</td>
</tr>
<tr>
<td></td>
<td>Hourly □</td>
</tr>
<tr>
<td></td>
<td>Daily □</td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
</tr>
</tbody>
</table>
| 2 | Yes □   | No □     | Does the employee ever receive feedback about their performance? If YES, indicate below:  
By whom? How often? How long after the task? Check all that apply: Feedback focus: □ Positive □ Corrective Feedback type: □ Written □ Verbal □ Grapheď □ Other |
| 3 | Yes □   | No □     | Does the employee ever see the effects of accurate task completion? If YES, how? (i.e., once the employee has completed the task as required, do they see the final product or are they provided with any reports about the effects of their work [e.g., happier clients, increased skills, reduction of client problem behavior]). If YES, how? |
| 4 | Yes □   | No □     | Is the task simple or does it involve relatively low response effort? |
| 5 | Yes □   | No □     | From your perspective as the supervisor, does the task generally take precedence over other potentially competing tasks? (i.e., has the employee ever told you that they could not complete the task because completing another task was a higher priority for them, or does the employee tend to choose to complete certain tasks over others when given a choice?). If NO, indicate these competing tasks below: Task 1 Task 2 Task 3 Task 4 |

**SCORING SUMMARY**

Rank the items that you scored as or each section as “NO”. Each item scored as “NO” is an opportunity for intervention. Prioritize the most problematic areas first.

1. [Blank]
2.
3.
4.
5.

Any other Concerns:
The table below outlines intervention planning strategies for different domains. Each item scored as 'NO' on the PDC-HS (1.1) should be considered as an opportunity for intervention with priority given to areas in which multiple items are endorsed. Interventions may be implemented concurrently or consecutively, with the latter option being preferred for settings in which staff resources are limited. Sample interventions and illustrative literature citations for each area are provided below.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Item #</th>
<th>Sample Intervention(s)</th>
<th>Literature Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td>1, 2, 3, 4, 5</td>
<td>Behavioral skills training (i.e., instructions, modeling, rehearsal, feedback)</td>
<td>Barnes, Dunning, &amp; Rehfeldt (2011); Parsons, Rollyson, &amp; Reid (2012)</td>
</tr>
<tr>
<td></td>
<td>1a</td>
<td>Improved personnel selection</td>
<td>Gatewood, Feild, &amp; Barrick (2016)</td>
</tr>
<tr>
<td></td>
<td>1b</td>
<td>Enhanced written instructions</td>
<td>Graff &amp; Karsten (2012)</td>
</tr>
<tr>
<td></td>
<td>1c, 1d</td>
<td>In-vivo/video modeling</td>
<td>Catania, Almeida, Liu-Constant, &amp; DiGennaro Reed (2009); Vladesou, Carroll, Paden, &amp; Kodak (2012)</td>
</tr>
<tr>
<td><strong>Task Clarification &amp; Prompting</strong></td>
<td>1, 2</td>
<td>Task clarification &amp; Checklists</td>
<td>Gravina, VanWagner, &amp; Austin (2008); Reetz, Whiting, &amp; Dixon (2016)</td>
</tr>
<tr>
<td></td>
<td>3, 4</td>
<td>Prompts</td>
<td>May, Austin, &amp; Dymond (2011); Loughrey, Marshall, Bellizzi, &amp; Wilder (2013); Downing, Capriola, &amp; Geller (2018)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Change/alter task location</td>
<td>Green, Reid, Passante, &amp; Canipe (2008)</td>
</tr>
<tr>
<td></td>
<td>2, 3, 4, 5</td>
<td>Improve access to (2, 3), redesign (4), or reorganize (5) task materials</td>
<td>Casella, Wilder, Neidert, Rey, Compton, &amp; Chong (2010); Abellon &amp; Wilder (2014)</td>
</tr>
<tr>
<td></td>
<td>6, 7</td>
<td>Reassess task process and personnel</td>
<td>Diener, McGee, &amp; Miguel (2009); McGee &amp; Diener (2010)</td>
</tr>
<tr>
<td><strong>Performance Consequences, Effort, &amp; Competition</strong></td>
<td>1</td>
<td>Increased supervisor presence</td>
<td>Bracket, Reid, &amp; Green (2007); Mowery, Miltenberger, &amp; Well (2010)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Performance feedback</td>
<td>Arco (2008); Green &amp; Dallery (2019); Rice, Austin, &amp; Gravina (2009)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Regularly highlight task outcomes</td>
<td>Methot, Williams, Cummings, &amp; Bradshaw (1996)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Reduce task effort</td>
<td>Abellon &amp; Wilder (2014); Casella et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Reduce aversive task properties</td>
<td>Green et al. (2008); Reed, DiGennaro Reed, Campisano, Lacourse, &amp; Azuly (2012)</td>
</tr>
</tbody>
</table>
REFERENCES


