THE ROLE OF ASSESSMENT IN IDENTIFYING EFFECTIVE
TEACHING INTERVENTIONS

BY

MEGAN HAFEN

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

———————————————————————
Chairperson Pamela L. Neidert, Ph.D.

———————————————————————
Florence D. DiGennaro Reed, Ph.D.

———————————————————————
Claudia L. Dozier, Ph.D.

———————————————————————
Debra Kamps, Ph.D.

———————————————————————
Barbara J. Thompson, Ph.D.

Date Defended: August 26, 2013

The dissertation committee for MEGAN HAFEN
certifies that this is the approved version of the following dissertation:

THE ROLE OF ASSESSMENT IN IDENTIFYING EFFECTIVE
TEACHING INTERVENTIONS

_____

Chairperson Pamela L. Neidert

Date approved: August 26, 2013

Abstract

Experimental analyses are designed to identify the variables maintaining responding, the results of which can be used to develop a treatment that directly addresses the function of the behavior. Experimental analyses of acquisition are a means to quickly compare treatment alternatives to identify the conditions that are likely to result in child learning. Assessment conditions are typically designed to identify skill versus performance deficits, and a number of variations in experimental arrangement have been reported. The purposes of the current study were to (a) replicate the results of previous research, specifically those obtained by Lerman et al. (2004), with a younger population with no known diagnoses and (b) compare three experimental designs in terms of efficiency and validity. The methodology designed by Lerman et al. was sufficient to identify an effective intervention for 20 of the 23 tasks that were assessed in Study 1. Results of Study 2 indicated that the brief multielement design was most efficient while the standard reversal was most efficacious. Given these findings, potential modifications to the assessment arrangement to enhance efficiency, while maintaining a high degree of predictive validity, are discussed.

Acknowledgements

First, I would like to acknowledge my advisor Pamela L. Neidert for her guidance and feedback in developing and refining this research project. Second, I would like to acknowledge the graduate students who contributed to the development, analysis, and conceptualization of these studies. Third, I would like to thank the members of my dissertation committee, Claudia L. Dozier, Florence D. DiGennaro Reed, Debra Kamps, and Barbara J. Thompson, for their feedback and assistance in further refining this project. Finally, I express sincere appreciation to our many undergraduate research assistants for their numerous contributions to these studies, with a special thanks to Paige J. Nguyen and Kristen N. Hastert for their thorough and diligent involvement with data collection, transfer, and analysis.

Table of Contents

The role of assessment in identifying effective teaching interventions

Students present with a variety of educational needs, highlighting the importance of

assessment practices in educational institutions. To enhance the utility of these practices, great

value can be obtained from assessment procedures that function not only to identify performance

problems but that also identify the specific needs of each student, inform the selection of

instructional interventions, and allow one to evaluate student progress within a pre-established

curriculum (Kratochwill & Sheridan, 1990; Noell, Ardoin, & Gansle, 2009; Shapiro & Derr,

1990). The field of behavior analysis has played an influential role in developing assessment

practices that establish a strong link between assessment and treatment. By analyzing antecedent

and consequent events and their corresponding influence on responding, behavior analysts can

(a) determine the variables controlling responding and (b) manipulate these variables to change

behavior. This approach to understanding and changing behavior led first to the development of

a methodology for assessing and treating problem behavior. Functional analyses of behavior

disorders involve exposing the individual to various experimental conditions in which the

experimenter alters the antecedent arrangement and arranges a potential source of reinforcement

(e.g., attention, escape) contingent on problem behavior. Levels of problem behavior in these

test conditions are compared to levels of problem behavior in a control condition in which

potential sources of reinforcement are provided noncontingently (or withheld following problem

behavior). The test conditions with the highest levels of problem behavior, relative to the control

condition, are indicative of the variables maintaining the aberrant behavior (Iwata & Dozier,

2008; Iwata, Dorsey, Slifer, Bauman, & Richman, 1982/1994). This information then is used to

prescribe treatment strategies that directly alter the behavior's maintaining variables (Hagopian,

Dozier, Rooker, & Jones, 2012; Vollmer, Iwata, Duncan, & Lerman, 1993; Vollmer & Northup,

1996). Researchers have demonstrated the utility of this assessment procedure by demonstrating (a) that function-based interventions are more effective than non-function-based interventions at decreasing levels of problem behavior (e.g., Newcomer & Lewis, 2004) and (b) that a variety of antecedent- and consequent-based interventions are effective at reducing levels of problem behavior when they directly address the behavior's controlling variables (e.g., Carr, Coriaty, & Dozier, 2000; Carr, Coriaty, Wilder et al., 2000; Tiger, Hanley, & Bruzek, 2008; Vollmer & Iwata, 1992).

This empirically supported approach of conducting experimental analyses to identify behavioral function that subsequently informs treatment selection has been extended from the assessment and treatment of problem behavior to the assessment and treatment of delayed learning and skill acquisition. Baer (2005) described the critical role of analysis in educational contexts. When teaching techniques are ineffective at producing the desired performance, responding must be analyzed to determine the function of poor performance. Baer provided a list of potential rationales for poor academic performance, including lack of reinforcer potency, ineffective prompting procedures, and a lack of necessary prerequisite skills. Studies utilizing functional analysis methodology to analyze the function of delayed skill acquisition include test conditions designed to address these rationales. Use of this methodology allows quick comparisons of treatment alternatives to determine the conditions under which appropriate behavior will and will not occur such that the results of the assessment can guide the selection of instructional interventions. This approach has been used to assess variables affecting oral reading performance (e.g., Daly, Martens, Dool, & Hintze, 1998; Eckert, Ardoin, Daisey, & Scarola, 2000; Noell, Freeland, Witt, & Gansle, 2001), reading comprehension (e.g., Lahey, McNees, & Brown, 1973), writing skills (e.g., Burns, Ganuza, & London, 2009), and correct

completion of academic work or curriculum goals (e.g., Duhon et al., 2004; Lerman, Vorndran, Addison, & Kuhn, 2004; McComas et al., 1996). These studies typically assessed performance under baseline arrangements in which motivational interventions and instructional prompts were absent. Subsequently, performance under these conditions was compared to performance under conditions in which one or more motivational or instructional procedures were introduced (Duhon et al., 2004). Such analyses attempted to form distinctions between performance (i.e., motivation) and skill deficits to highlight the types of instructional interventions that were likely to yield optimal performance under more typical teaching arrangements.

Precedent and Procedures for Identifying Performance Deficits. The term *performance deficit* is used when the presence or absence of motivational strategies is sufficient to influence responding (Elliott & Shapiro, 1990; Jones & Wickstrom, 2002). Identification of performance deficits would suggest that the use of reinforcement contingencies or other motivational strategies (e.g., interspersing known tasks with unknown tasks, incorporating opportunities to choose tasks or reinforcers, implementing error correction for incorrect responding) should be sufficient to occasion correct responding.

This logic is supported by early studies, which have demonstrated improved performance as a function of motivational interventions. For example, Ayllon and Kelly (1972) and Ayllon and Roberts (1974) demonstrated improvements in academic performance after incorporating reinforcement contingencies for accurate responding. Ayllon and Kelly presented a standardized test, as outlined in the test's administration manual (i.e., baseline procedures), to 12 participants with developmental disabilities. That same day, the test was readministered. After each subsection of the test booklet was completed, the participants' test books were scored. The participants earned a token, exchangeable for back-up reinforcers, for each of their correct

3

answers. Significant increases in test performance were observed under the reinforcement administration as compared to the standard baseline administration. In other words, test results were higher during a condition in which preferred items were arranged for correct responding as compared to a condition in which no preferred items were provided. These data suggest that contingent access to preferred items may be sufficient to influence academic performance by enhancing the establishing operation for correct responding.

Improvements in academic performance have also been observed by interspersing known (i.e., maintenance) tasks with unknown (i.e., acquisition/target) tasks (Koegel & Koegel, 1986; Neef, Iwata, & Page, 1980). Koegel and Koegel (1986) compared the number of correct, unprompted responses during an acquisition-only phase to a phase in which acquisition and maintenance tasks were interspersed. Increases in correct responding, ranging from approximately 20-35%, were observed during the interspersal phase for three of the four academic areas, suggesting that in some situations, simply interspersing known tasks with unknown tasks may be sufficient to increase correct responding. Researchers have suggested that the behavioral mechanisms responsible for improved performance under interspersal conditions may be attributed to (a) an increase in the overall rate of reinforcement (Koegel & Koegel, 1986), (b) establishing a motivating operation by enhancing the value of reinforcement for correct responding (Volkert, Lerman, Trosclair, Addison, & Kodak, 2008; Neef et al., 1980) or (c) enhancing attending to acquisition tasks as a function of task variation (Dunlap & Koegel, 1980; Koegel & Koegel, 1986) or an established momentum of compliance (Mace & Belfiore, 1990).

A third motivational strategy that may influence performance is choice making. Dunlap et al. (1994) observed increases in task engagement and decreases in disruption when

participants were provided a choice of academic assignment as compared to a condition in which the teacher selected assignments. Similarly, Tasky, Rudrud, Schulze, and Rapp (2008) observed increases in task engagement when participants were allowed to select the household chores they completed. Increases in task engagement were observed even when this same sequence of chores was presented in a yoked-control condition. Other researchers (e.g., Fenerty & Tiger, 2010; Fisher, Thompson, Piazza, Crosland, & Gotjen, 1997; Schmidt, Hanley, & Layer, 2009) have demonstrated a preference for choosing reinforcers even when reinforcement was equated across choice and no-choice conditions (i.e., the same stimulus was delivered across conditions regardless of whether selection was made by the participant or therapist). Researchers have speculated that control over reinforcement may function as (a) a reinforcer (Fisher et al., 1997) or (b) an establishing operation, enhancing the value of the stimulus presented (Romaniuk et al., 2002).

A fourth motivational strategy for improving academic performance is the use of error correction. Although error correction may be seen as an instructional strategy that functions to enhance stimulus control over correct responding by incorporating additional opportunities for the correct response to occur under the appropriate stimulus conditions (Worsdell et al., 2005), several researchers (Cuvo, Ashley, Marso, Zhang, & Fry, 1995; Rodgers & Iwata, 1991; Worsdell et al., 2005) have demonstrated increases in correct responding even when the correction trial did not provide practice of the target response (i.e., the student was asked to emit a response different from the one required on that learning trial). These improvements in performance during irrelevant error correction procedures suggest that error correction functions as punishment for incorrect responding. In other words, one can avoid the error correction

procedure by responding correctly to the target task, suggesting that error correction influences correct responding as a function of negative reinforcement.

Together, research in this area suggests that motivational strategies alone may be sufficient to increase academic performance. Thus, motivational strategies are typically incorporated into academic skill assessments to evaluate their effect on responding. The use of criterion-contingent rewards has been most commonly utilized (e.g., Bonfiglio, Daly, Martens, Lin, & Corsaut, 2004; Daly, Martens, Hamler, Dool, & Eckert, 1999; Daly, Murdoch, Lillenstein, Webber, & Lentz, 2002; Wagner, McComas, Bollman, & Holton, 2006) to assess potential performance deficits. With this procedure, experimenters examined the participants' baseline level of responding and used this information to establish a criterion for reinforcement. If the participant met, or in some cases exceeded, the predetermined criterion, the participant was allowed to select a reward (typically a tangible or edible item).

Ultimately, the identification of performance deficits aids in treatment selection. After identifying the motivational strategies that are effective at increasing academic performance, the goal of intervention then could be to thin the schedule of reinforcement or decrease the ratio of known to unknown tasks, for example. Thus, rather than introducing irrelevant and potentially time consuming prompting and prompt-fading strategies, the teacher could simply incorporate the effective motivational strategy into the teaching context and gradually thin its usage such that the participant continues to respond correctly even as use of the motivational strategy is thinned.

Precedent and Procedures for Identifying Skill Deficits. When motivational strategies are insufficient to increase academic performance, the effectiveness of prompting or other instructional strategies is typically examined. This is done to determine whether poor academic performance is due to a skill deficit. The term *skill deficit* describes responses that have not been

previously acquired and will fail to occur frequently and consistently in the absence of instructional interventions (Daly et al., 1998; Duhon et al., 2004; Lerman et al., 2004).

Several brief assessments have been conducted that exclusively examined the influence of various prompting procedures on task completion (McComas et al., 1996; Richman et al., 2001). McComas et al. (1996), for example, were interested in identifying prompting strategies (e.g., providing a written model of a rhyming word, providing a verbal outline of a reading passage) that were effective at increasing accurate spelling or reading comprehension for four participants. A number of strategies were evaluated sequentially, beginning with interventions that required the least amount of adult assistance. McComas et al. demonstrated systematic patterns of responding when the effective prompting strategy was both added and removed, suggesting that prompts alone may be sufficient to increase correct responding but that all prompting strategies may not be equally effective at producing the desired performance for a given individual.

Other researchers (e.g., Cuvo, Davis, O'Reilly, Mooney, & Crowley, 1992; Roll, 1973) have evaluated the effectiveness of feedback for enhancing performance. Cuvo et al. (1992) demonstrated that the combination of textual prompts and performance feedback was more effective for some participants than textual prompts alone for correctly completing community living skills. Roll (1973) demonstrated that performance feedback, in the form of colored lights that signaled nasalized and non-nasalized vocalizations, was effective at reducing the production of nasalized phonemes for two children with cleft palates. Studies such as these suggest that feedback may be effective at influencing performance; however, feedback is often implemented as part of a package intervention (Cuvo et al., 1992; Fueyo & Bushell, 1998; Kern-Dunlap et al., 1992; Odom, Chandler, Ostrosky, McConnell, & Reaney, 1992; Schulman, Suran, Stevens, &

Kupst, 1979) and is often evaluated in combination with motivational interventions (Kern-Dunlap et al., 1992; Schulman et al., 1979). Thus, the effectiveness of this procedure in isolation is relatively unclear.

Given research suggesting that instructional strategies may be effective at influencing academic performance, skill assessments typically include conditions that evaluate the effects of one or a few instructional strategies on academic responding. In contrast to the conditions used to identify performance deficits, in which a given procedure (i.e., criterion-contingent rewards) was common across studies, various assessment conditions have been designed to test for skill deficits, often utilizing prompting procedures that are idiosyncratic to the dependent variable. For example, Burns, Ganuza, and London (2009) placed graph paper beneath a sheet of writing paper and taped it to the student's desk. This served as a visual prompt in an effort to enhance their participant's correct written-letter formation. In another example, Duhon et al. (2004) provided their participant with a table of 3-digit multiplication facts in an effort to improve their participant's completion of 3-digit-by-3-digit multiplication problems. Use of the table for solving problems was modeled prior to each session. Despite variations in the form of the instructional strategy that are implemented across studies, the implications given improved performance are the same. Ultimately, if the assessment data suggest that instructional strategies alone are sufficient to increase correct responding, this information can be used to develop an effective classroom treatment. Ideally, the instructional strategy (e.g., the effective prompting procedure) would be introduced into the classroom with the goal of systematically fading its use (e.g., fading the delivery of the prompt).

When instructional or motivational strategies alone are insufficient to increase academic performance, the effectiveness of these strategies in combination is often evaluated. In addition

to evaluating the combined effectiveness of these strategies, this condition rules out the possibility that the chosen strategies simply are ineffective at improving performance on a given skill (Lerman et al., 2004). More specifically, if increases in correct responding are not observed under the combined condition, these data would suggest that alternative instructional or motivational strategies should be assessed or that the student may lack necessary prerequisite skills, suggesting that the assessment be conducted again using easier materials or targeting alternative skills. Alternatively, if responding maintains at high levels during the combined condition, these data suggest that the combination of instructional and motivational strategies are necessary for maintaining high levels of performance. This finding highlights the need to systematically fade both components of the intervention (e.g., thinning the schedule of reinforcement and fading the delivery of prompts).

Variations in Experimental Design. Although the use of assessment conditions to identify skill and performance deficits is common throughout the skill assessment literature, a number of variations in experimental design have been reported. Those most commonly used involve evaluations of responding during a brief multielement assessment, a standard multielement assessment, or an assessment procedure in which repeated measures of responding are obtained during sequential presentation of the independent variables.

Brief Multielement Design. The brief multielement assessment involves a single evaluation (i.e., one data point per condition) of responding across experimental conditions in a sequential order. That is, conditions are presented in a predetermined sequence, based most commonly on the intrusiveness of the intervention or its ease of implementation. This is done in an attempt to identify the instructional intervention that requires the least amount of adult involvement but produces the most discriminable improvement in performance as compared to

baseline (Daly, Witt, Martens, & Dool, 1997; Martens, Eckert, Bradley, & Ardoin, 1999).

Evaluations of responding under the various experimental conditions continue until an

intervention is identified that produces a noticeable increase in academic responding. At that

point, a brief reversal is conducted.

Given that repeated measures of responding under each of the experimental conditions

are lacking, trend and variability analyses are impossible. Thus, this assessment approach uses

changes in level as the indicator of treatment effectiveness. If the treatment in question produces

consistently higher levels of performance than baseline or the preceding treatment, the brief

reversal provides some confidence that the change in behavior is the result of changes in the

independent variable (Daly et al., 1997; Martens et al., 1999). Some researchers (e.g., Burns et

al., 2009; Gortmaker, Daly, McCurdy, Persampieri, & Hergenrader, 2007; Jones & Wickstrom,

2002) have incorporated a phase of extended treatment implementation following the assessment

as a means to further validate the assessment results.

A number of studies (Daly et al., 1998; Daly et al., 1999; Gortmaker et al., 2007; Jones,

& Wickstrom, 2002; Noell et al., 2001) suggest that the brief multielement assessment has the

potential to identify effective interventions; however, a priori decisions should not be made

regarding the number of sessions conducted during the brief assessment. That is, visual

inspection of the data should guide assessment length. In some situations, a single replication of

the observed effect may provide sufficient demonstration of treatment effectiveness, but in other

situations, multiple replications may be useful or necessary to adequately identify a promising

intervention, especially when trends or marked variability are present.

Additionally, the brief multielement assessment may provide the most clear and useful

results when assessment conditions are carefully designed to address each potential deficit. In

this process, it may be worthwhile to combine a number of potentially effective interventions into a single assessment condition rather than evaluating a large number of motivational or instructional strategies in isolation (Lerman et al., 2004). After identifying whether poor responding is due to a skill or performance deficit, further analyses then could be conducted to identify a specific function-based intervention that is least intrusive and most efficacious.

Standard Multielement Design. An alternative to using the brief multielement assessment is the use of a standard multielement assessment in which test (i.e., treatment) and control (i.e., baseline) conditions are alternated in rapid succession over the course of several series. Differentiated responding between test and control conditions is indicative of a treatment effect. Comparisons between treatment procedures can be made by analyzing the amount of differentiation (a) between test and control conditions and (b) between different test conditions. Because the standard multielement assessment involves repeated measures of responding under each of the assessment conditions, level, trend, and variability analyses can be conducted, providing more detailed information regarding the differential effectiveness of various teaching interventions.

A number of researchers (Daly, Bonfiglio, Mattson, Persampieri, & Foreman-Yates, 2005; Eckert et al., 2000; Eckert, Ardoin, Daly, & Martens, 2002; Wagner et al., 2006) have demonstrated the potential for differentiated patterns of responding using this experimental design. However, given that multiple conditions are conducted simultaneously and discriminations are made across data paths, the clarity of the assessment results may be influenced by the number of assessment conditions evaluated simultaneously (e.g., Eckert et al., 2000; Eckert et al., 2002). By assessing fewer conditions, one may decrease the likelihood of discrimination failure, increase visual clarity of the assessment results by minimizing the number

of data paths for which conclusions must be drawn, and enhance the efficiency of the assessment procedure. In addition, rather than assessing all potential treatment combinations, the experimenter may identify the least intrusive assessment conditions that aim to identify skill and performance deficits. If increases in performance relative to baseline are not observed, additional assessment conditions then may be evaluated. Alternatively, experimenters may consider evaluating a number of assessment conditions initially. As ineffective strategies are identified, these conditions could be discontinued from the analysis while the remaining conditions continue to be evaluated. At this point, additional conditions then could be introduced such that information is obtained regarding the treatment efficacy of a number of interventions while minimizing the number of discriminations that must be made at any given point in time (e.g., Wagner et al., 2006).

Sequential Presentation of Assessment Conditions. A third type of experimental assessment foregoes some efficiency to obtain repeated measures of responding under each assessment condition. With this approach, independent variables are introduced sequentially, and repeated measures of responding are obtained during each condition, providing some information as to the effects of repeated exposure on responding and the extent to which responding maintains over time. Several experimental designs have been used to demonstrate control over responding. Noell et al. (1998) examined the number of words read correctly per min (WCPM) across three instructional levels using a multiple baseline design. When easier material was presented, two of the three participants performed at high levels during a baseline or continent reward condition. When more advanced reading material was presented, modeling and practice were necessary. To demonstrate experimental control under this arrangement, systematic increases in WCPM would need to be observed as a given intervention is introduced

across each baseline. However, because different interventions were effective across different instructional levels, experimental control was lacking. Future researchers may need to stagger implementation across more than three panels in order to demonstrate sufficient replication of the desired effect. Alternatively, future researchers may incorporate reversals in which the effective intervention is removed and subsequently reintroduced in order to enhance the experimental rigor.

A second means of experimental control would involve a multielement evaluation of matched and mismatched interventions based on the assessment results. A somewhat similar approach was used by Duhon et al. (2004) to validate the results of their initial assessment. For all four participants, the authors compared responding during a skill-based intervention to that of a performance-based intervention. Differentiation between these data paths, in the direction predicted by their initial assessment, was observed for all four participants. Although Duhon et al. used this arrangement as a more extended validation procedure, this arrangement could be used in isolation to identify the function of delayed acquisition. To do this, assessment conditions could be presented sequentially, and repeated measures of responding could be obtained under each experimental condition. Subsequently, matched and mismatched interventions could be selected from these data and evaluated in a multielement design as a means of experimental control.

The use of a reversal design is a third means of experimental control that has been utilized. Lerman, Vorndran, Addison, and Kuhn (2004) used this procedure to identify assessment conditions that would produce consistent increases in early learner skills (e.g., matching, listener responding, play skills, fine motor skills, preacademic skills) for six children with autism. First, the authors identified two to three target tasks for each participant. In order

13

to examine the differential effectiveness of various teaching strategies, it was necessary to identify tasks for which improvements in performance were necessary. Thus, target tasks were identified by examining the level of correct responding in the absence of instructional or motivational strategies (i.e., baseline conditions). In these baseline sessions, 10 trials of the task were presented and no consequences were arranged for correct or incorrect responding. Following baseline, a reinforcement condition was implemented in which maintenance trials were interspersed with the 10 target trials. Additionally, choice-making opportunities (e.g., choice of preferred item or session materials) were incorporated, and praise and preferred items were delivered contingent on correct responding. If responding failed to improve under this condition, a prompts condition was implemented in which a combination of response and stimulus prompts was delivered simultaneous with the target instruction. If correct responding did not increase or failed to maintain under these conditions, a combined condition was implemented that contained all of the reinforcement and prompting procedures described previously. After identifying an effective intervention, a reversal was conducted. Across all target tasks for each of the participants, the authors observed that one of the assessment conditions was consistently effective at increasing levels of correct responding relative to baseline or the other assessment conditions. Additionally, for 15 of the 16 tasks, terminal responding occurred at clinically significant levels.

This study was of great importance to the skill assessment literature. Lerman et al. (2004) were one of the first to develop a set of general procedures that could be used to systematically assess performance on a variety of pre-academic tasks. The authors designed assessment conditions to test broadly for potential deficits, by combining a number of interventions into each assessment condition, rather than evaluating a number of specific

motivational or instructional interventions in isolation. This approach had rarely been used in

the skill assessment literature and was beneficial for increasing the efficiency of the assessment

by minimizing the likelihood of idiosyncratic responding to specific independent variables.

Additionally, this broad assessment approach allowed extension beyond oral reading fluency in

school-age participants. Few researchers prior to this time (Lahey et al., 1973; McComas et al.,

1996) assessed the generality of this methodology when extended to other skill domains or to

younger populations of participants, with or without intellectual and developmental disabilities.

Ultimately, the data obtained by Lerman et al. (2004) suggested that their methodology

was a useful way to assess academic performance in a clear manner. Although idiosyncratic

results were observed across skills and participants, systematic changes in performance were

observed for each of the target tasks. However, in contrast to the skill assessment procedures

described previously (i.e., brief and standard multielement), one of the biggest limitations of

their study was arguably the duration of time that was required to complete the assessment.

Although sessions were only 3 to 10 min in duration, between 12 and 40 sessions were required

to complete the assessment for each target task. Thus, despite their clear and promising results,

the amount of time required to complete the assessment may prove overly demanding,

particularly in contrast to the assessment methodologies described previously. Research

examining the usefulness of this procedure using a more time efficient experimental arrangement

(i.e., brief or standard multielement) is warranted.

This need is consistent with gaps in the existing literature, suggesting that a useful

direction for future research is to compare the treatment utility of these different assessment

approaches (Eckert et al., 2000; Daly et al., 2005; Daly et al., 1998; Daly et al., 1999). To

accurately assess treatment utility, research is needed to demonstrate that the interventions

identified by these assessments prove effective over extended implementations. Comparisons across the three assessment approaches would permit conclusions regarding the efficiency and accuracy of each approach and may be useful in developing an assessment progression that addresses issues with both efficiency and utility. Additionally, although Lerman et al. (2004) played an important role in extending skill assessment methodology to novel target behaviors, further research is needed to demonstrate the usefulness of these experimental analyses across a range of educational concerns (Duhon et al., 2004; Lerman et al., 2004; Noell, Roane, VanDerHeyden, Whitmarsh, & Gatti, 2000; Noell et al., 2001). Similarly, the vast majority of previous research has targeted elementary-aged children, particularly between 1st and 4th grades (e.g., Bonfiglio et al., 2004; Daly et al., 1998; Daly et al., 1999; Duhon et al., 2004; Eckert et al., 2000; Gortmaker et al., 2007; Noell et al., 2001; Wagner et al., 2006). As such, the extension of these procedures to a younger population of participants is warranted. In sum, despite the large amount of research in this area suggesting that the use of experimental analyses to address educational concerns has substantial merit, additional research would be valuable in refining and extending this methodology.

Purpose

Given that Lerman et al. (2004) were one of the first to develop a set of general procedures designed to test broadly for potential deficits, the purpose of the present study was to replicate the effectiveness of their procedures with a younger population of participants with no known diagnoses. Subsequently, an examination was conducted to compare the efficiency and validity of two common skill assessment methodologies (i.e., brief multielement, standard multielement) with the standard reversal used by Lerman et al.

Method: Study 1

16

Participants and Setting

Participants were 8 toddlers with no known diagnoses, ranging in age from 16 to 26 months. One participant (Zara) was referred by her doctor to an early intervention specialist for concerns regarding speech and language development. Results of a comprehensive speech and language evaluation suggested that Zara had a limited phonemic repertoire and decreased ability to imitate words; however, her language-comprehension abilities were at and above grade level. All other aspects of Zara's development were occurring in a typical developmental progression. All participants attended a university-based daycare classroom and had mastered a number of skills in their classroom curriculum prior to participation in this study. Detailed information (e.g., ages, mastered skills) regarding each of the participants can be found in Table 1.

Sessions were conducted in a segmented area of the children's classroom, which contained two tables, chairs, and all necessary sessions materials. This area was partitioned from the rest of the classroom by a 2-ft barrier. During sessions, the participant was seated with his/her back to the classroom; however, ongoing activities could be seen and heard with minimal effort.

Response Measurement and Reliability

Data were collected on the frequency of target instructions and correct responding. Target instructions were defined as experimenter presentation of target-task materials paired with a relevant vocal instruction (e.g., "Sort the pieces"). Correct responding was defined as child initiation of the requested action within 5 s of the experimenter's instruction and completion of the desired response within 10 s of the instruction. Given that the assessment procedure was designed to examine the influence of various teaching strategies (i.e., motivational interventions or prompting procedures) on responding, correct responding that occurred following the

experimenter prompt (in phases in which prompts were delivered) was scored as correct. This was done to provide information on the differential influence of prompts on the participants' academic performance as compared to conditions in which no prompting procedures were arranged.

The dependent variable was the percentage of trials with correct responding to the target task. Ten trials of the target task were presented in each session. Thus, to calculate the percentage of trials with correct responding, the number of correct responses were divided by 10 and multiplied by 100%. Data were also collected on the frequency with which maintenance instructions, prompts, praise, and preferred items were delivered such that measures of treatment integrity could be calculated. Maintenance instructions were defined as experimenter presentation of maintenance-task materials paired with a relevant vocal instruction (e.g., "Stack the rings"). Prompts were defined as the experimenter modeling the desired response and altering the presentation of task materials (e.g., placing the correct stimulus closer to the participant than the distracter stimuli) to occasion the desired response. Praise was defined as brief statements of approval or commendation delivered by the experimenter contingent on task responding. The delivery of preferred items was defined as experimenter presentation of edible or leisure items.

A second observer simultaneously, but independently, collected data during at least 37% of the sessions for each participant. Agreement scores were determined by calculating occurrence or nonoccurrence agreement within trials for each of the dependent variables. Trial initiation was indicated by the delivery of a target or maintenance instruction. Trial termination was indicated by the last scored response (e.g., prompt, correct responding, praise, delivery of preferred items) that immediately preceded the delivery of a subsequent target or maintenance

instruction. Within each trial, observers compared agreements and disagreements for each of the

dependent variables. To calculate interobserver agreement, the number of trials with agreements

was divided by the total number of trials and multiplied by 100%. Mean interobserver

agreement for target instructions and correct responding was 96% (range, 18% to 100%) and

95% (range, 29% to 100%), respectively. Mean interobserver agreement for the delivery of

maintenance instructions, prompts, praise, and preferred items was 97% (range, 18% to 100%),

98% (range, 38% to 100%), 96% (range, 29% to 100%), and 97% (range, 54% to 100%),

respectively.

Treatment integrity percentages were calculated for each session. Across the four

experimental conditions (i.e., baseline, motivational interventions, prompts, and combined), 12

potential errors of omission or commission were calculated. To obtain an overall score, we

summed the number of procedural steps implemented correctly (across phases, tasks, and

participants), divided by the total number of procedural steps available, and multiplied by 100%.

Treatment fidelity for Study 1 was 97%.

Procedure

A systematic replication of the procedures described by Lerman et al. (2004) was

conducted. Tasks for each participant were selected from the participant's classroom curriculum

or popular curriculum guides (e.g., Partington, 2006; Sundberg, 2008). Tasks were excluded for

use in the study if (a) skill mastery had been documented or (b) the task was one of the learning

objectives targeted for intensive teaching in the classroom. The selected tasks targeted the skill

domains of listener responding (e.g., receptive identification of colors, letters, animals, objects),

visual perceptual skills and matching-to-sample (e.g., sorting), imitation (e.g., building identical

block structures), independent play (e.g., completing puzzles or a shape sorter), or fine motor

development (e.g., twisting washer on plastic screw). Before conducting sessions, a 10-item paired-choice preference assessment was conducted, using procedures similar to those described by Fisher et al. (1992), to identify highly preferred edible or leisure items for each participant. Preference assessment items were selected at random by the experimenter. After completing the assessment, the top three ranked items were identified by calculating the percentage of trials in which each item was selected. If selection percentages were tied across several items, all items with the same percentage of selection as the top three items were included in Study 1. Selected items were unavailable to the participant in the classroom outside of experimental sessions, with a few exceptions. First, if the child was simultaneously participating in an intensive toilet-training regimen in their classroom, some of the same items may have been used as reinforcers in that setting. Second, bubbles for Zara and vanilla wafers for Brandy were available in the classroom on occasion.

In all conditions, the participant was seated at a table with the experimenter and data collector(s). Prior to each session, the experimenter modeled the correct response(s). If the target task involved multiple discriminations (e.g., identifying animals in an array of three), the experimenter modeled the correct response to each stimulus. The experimenter then physically guided the participant to complete the response(s). The purpose of this pre-session prompting was to provide contact with the appropriate response requirement before any given session. Given the age of our participants, the children may have lacked a history of task completion in the presence of specific vocal instructions (e.g., "Match this"). This pre-session prompting helped mitigate any issues with limited listener-responding repertoires. Following this pre-session prompting, the instructional trials were presented. Instructional trials involved (a) presentation of the task materials and the delivery of a vocal instruction (with or without

prompts), (b) the participant's response (or lack thereof) to the instruction, and (c) the experimenter's response (or lack thereof) to the participant's behavior. Sessions continued until 10 trials of the target instruction were presented. All disruptive behavior was ignored. Typically, two to seven sessions were conducted per day, four to five days per week.

Baseline. At the start of each trial, the task materials were presented and the experimenter delivered a relevant instruction (e.g., "Finish the puzzle"). If the participant made a correct response within 5 s (or 5 s elapsed with no responding), the materials were removed, and the trial was terminated. If the participant initiated a response (touched the materials) within 5 s but did not complete the task, an additional 5 s was provided to complete the requested action. When 10 s had elapsed (regardless of whether the participant had completed the response), the materials were removed, and the trial was terminated. No programmed consequences were delivered for correct or incorrect responding.

The purpose of this phase was (a) to determine the level of correct responding in the absence of prompting and motivational strategies and (b) to identify two to three target tasks and a maintenance task (to be used in the motivational interventions and combined conditions) for each participant. Target tasks were defined as those to which the mean level of correct responding fell at or below 50%. For tasks involving multiple discriminations (e.g., identifying animals in an array of three), the data were analyzed further to determine whether correct responding occurred consistently to any of the stimuli in the array (e.g., the participant consistently identifies *bear* but not *owl* or *bee*). If so, this discrimination was removed from the array, replaced with an alternative discrimination, and baseline sessions with the modified task were conducted anew. Maintenance tasks were defined as those to which correct responding occurred at or above 80% of trials for two consecutive sessions or two of three consecutive

sessions, with the level of correct responding remaining at or above 70% for the third session. These maintenance tasks were used during the subsequent motivational interventions condition (see details below); thus, no target tasks were exposed to the motivational interventions condition until a maintenance task had been identified.

Motivational Interventions. Immediately preceding session, the participant's most preferred edible or leisure items (a minimum of three items that ranked highest on the participant's paired-choice preference assessment) were presented in an array in front of the participant. For six of the eight participants (Beth, Liv, Jocelyn, Ivy, Eric, and Leanne), edible items were included in the assessment. Leisure items were included for Zara per parent request. For Brandy, both leisure and edible items were included in the assessment but were evaluated separately. After presenting the array of preferred items, the experimenter instructed the participant to select the item for which they wanted to work. Following selection, pre-session prompting was provided, and then, the instructional trials were presented as described in baseline.

On any trial in which the participant responded correctly to the target instruction, praise and a preferred item were delivered (small piece of food or 15-s access to the selected toy). Maintenance instructions were interspersed with the target instructions in an effort to enhance attending to the target task and strengthen the establishing operation for correct responding. Thus, for the first few trials in which the maintenance instruction was presented, praise and a preferred item were also delivered for correct responding. This was done to increase the likelihood that the participant would contact the reinforcement contingency at least once per session. Correct responding to all subsequent maintenance instructions resulted exclusively in praise. This was done to ensure that the richer schedule of reinforcement favored the acquisition

task. Initially, the maintenance task was interspersed with the target task on a one-to-one basis (i.e., a maintenance task was presented after each target task). If correct responding began to increase under the motivational interventions condition, the maintenance task was presented after every two or three target tasks. This was done to decrease session duration and eliminate superfluous instructions if responding would maintain at high levels under a leaner interspersal schedule. These changes in the interspersal ratio were manipulated across sessions rather than within session. The purpose of this condition was to determine whether the introduction of motivational procedures would be sufficient to increase levels of correct responding.

Prompts. Prompting sessions were identical to baseline except response (i.e., modeling the correct response) and stimulus (i.e., altering the presentation of task materials) prompts were paired with the delivery of each target instruction. Stimulus prompts included (a) placing the correct stimulus closer to the participant than the other stimuli in the array (for listener-responding tasks), (b) positioning pieces slightly over their appropriate openings in the shape sorter or puzzle, (c) placing sorting materials directly in front of their respective location, (d) twisting the washer on the top portion of the plastic screw, or (e) arranging the first two blocks of the three-block structure. No programmed consequences were arranged for correct or incorrect responding. The purpose of this condition was to determine whether prompting strategies alone would be sufficient to increase correct responding.

Combined. Combined sessions were identical to the combination of procedures described in the motivational interventions and prompts conditions. More specifically, (a) response and stimulus prompts were paired with the delivery of each target instruction, (b) praise and a preferred item were delivered contingent on correct responding, (c) participants selected the preferred item for which they wanted to work, and (d) a maintenance task was interspersed

23

with the target task. The purpose of this condition was to determine whether the combination of prompts and motivational strategies would be sufficient to increase correct responding.

Data Analysis and Experimental Design

Data were analyzed by comparing the percentage of trials with correct responding across the experimental conditions using a reversal design. We used a reversal design, consistent with Lerman et al. (2004), because it provided the most conservative evaluation of performance, incorporating repeated measures of responding and replications of the desired effect. Following baseline, the motivational interventions condition was always implemented first. This was done to eliminate a seemingly superfluous evaluation of prompting procedures if a condition could be identified that would produce independent responding.

Data were collected in this condition until stable responding was observed. If no increase in correct responding was observed or the mean level of correct responding fell below 65%, the prompts condition was introduced. If the mean level of correct responding during the motivational interventions condition fell at or above 65%, a replication of the observed effect was conducted. That is, baseline conditions were re-implemented followed by a reintroduction of the motivational interventions condition. An exception to this rule was made for Eric when a sharp increasing trend was observed during the motivational interventions phase for his puzzle task. Given this, a replication of the observed effect was conducted.

If the mean level of correct responding in the subsequent motivational interventions phase fell below 80%, the prompts condition was introduced. This was done because 80% was selected as our criteria for clinical significance. If repeated exposure to the motivational interventions condition failed to reach this level, the data suggested that additional modifications were warranted in order to produce a clinically significant change in responding.

If the prompts condition produced a sizeable increase in the level of correct responding (i.e., the mean level of correct responding in the prompts condition fell at or above 65%), a replication of the observed effect was conducted in which the last ineffective intervention (i.e., motivational interventions) was presented followed by the reintroduction of the prompts condition. An exception to this rule was made for Ivy's receptive-identification-of-objects task in which the mean level of correct responding fell just below 65%, but a change in level was observed relative to the previous phases. Given this, a replication of the observed effect was conducted.

If a sizeable increase in the level of correct responding was not observed or the mean level of correct responding in the subsequent prompts phase fell below 80% (i.e., clinically significant levels of correct responding were not observed given repeated exposure to the assessment condition), the combined condition was introduced. Exceptions to this rule were made for Brandy's animal-identification tasks. Although the mean level of correct responding in the final prompting phases fell below 80%, a small number of sessions were conducted in each phase. In the top panel (Figure 2), the last three data points fell at 80%. In the bottom panel, an increasing trend was observed throughout the phase with the percentage of correct responding in the last two sessions occurring at 80% and 90%, respectively. Given this, prompting alone was identified to be an effective intervention and the analysis was considered complete. For the remaining participants, if further increases in correct responding were observed under the combined condition, a replication was conducted in which levels of correct responding in the combined condition were systematically compared to levels of correct responding in the motivational interventions and prompts conditions. If consistent increases in correct responding

were not observed across any of the conditions, as is the case with Zara's puzzle and shape sorter tasks, the analysis was considered complete and no further evaluations were conducted.

Results: Study 1

Results across target tasks are depicted for each participant in Figures 1 through 8. Sessions are depicted along the x-axis and the percentage of trials with correct responses along the y-axis. For Beth (Figure 1), an immediate and sustained increase in the level of correct responding was observed when motivational strategies were implemented for her sorting task. This effect was replicated following a return to baseline. For her puzzle task, increases in correct responding were observed during the motivational interventions and prompts conditions relative to baseline; however, clinically significant increases were not observed during either condition. When the combined condition was introduced, perfect performance was observed for three consecutive sessions. The effectiveness of the combined condition was replicated following returns to both the motivational interventions and prompts conditions. Finally, for her letter-identification task, the highest levels of correct responding were observed during the prompts condition. This effect was replicated following a return to the motivational interventions condition.

For both of Brandy's animal-identification tasks (Figure 2), the prompts condition resulted in the highest levels of correct responding. This effect was replicated following returns to the motivational interventions condition.

For Eric (Figure 3), a gradual increase in the level of correct responding was observed for his puzzle task during the initial implementation of the motivational interventions condition. Immediate increases in correct responding were observed during subsequent implementations of this condition relative to the level of responding in baseline. For his shape sorter task, a gradual

increase in correct responding was also observed during the initial exposure to motivational strategies; however, responding failed to reach clinically significant levels. A decrease in responding was observed when the prompts condition was implemented. We replicated this effect, but clinically significant levels of responding had yet to be obtained. Given this, the combined phase was introduced. High and stable levels of correct responding were observed, and this effect was replicated following returns to the motivational interventions and prompts conditions. For Eric's sorting task, near perfect performance was observed when the combined phase was introduced. This effect was replicated following a return to the prompts condition; however, levels of responding did not decrease to the levels observed previously when the motivational interventions phase was reintroduced. These data suggest that following a history of exposure to the combined condition, motivational strategies alone were sufficient to maintain relatively high levels of correct responding.

For Ivy (Figure 4), the combined condition produced the highest level of correct responding for her animal-identification and puzzle tasks. This pattern of responding was observed each time the combined phase was introduced following returns to the motivational interventions or prompts conditions. For Ivy's receptive-identification-of-objects task, the prompts condition was sufficient to increase correct responding, and this effect was replicated following a return to the motivational interventions condition.

For Jocelyn (Figure 5), the highest levels of correct responding to the shape sorter task were observed during the combined phase. This effect was replicated following returns to the motivational interventions and prompts conditions. For her receptive-identification-of-letters task, the prompts condition produced the highest level of correct responding, and this effect was replicated following a return to the motivational interventions phase. For her puzzle task,

increases in correct responding were observed when motivational strategies were introduced. A replication of this effect was observed following a return to baseline.

For Leanne (Figure 6), none of the conditions produced a clinically significant increase in Leanne's correct completion of the puzzle task. Given that the prompts condition involved fewer teaching components than the combined condition but produced a similar increase in correct responding, a reversal was conducted in an attempt to replicate this effect. During the second implementation of the prompts phase, a decrease in the level of correct responding was observed relative to the previous combined phase. When the combined phase was reintroduced, a gradual increase in correct responding was observed until responding reached a clinically significant level. At that point, the effectiveness of the combined condition was replicated following returns to the motivational interventions and prompts conditions. For Leanne's animal-identification task, an increase in correct responding was observed during the prompts condition; however, this effect was not replicated following a return to the motivational interventions condition. Given this, the combined phase was introduced, and high and stable levels of correct responding were observed. The effectiveness of the combined condition was replicated following returns to the motivational interventions and prompts conditions. Finally, for Leanne's receptive-identification-of-objects task, an increase in correct responding was observed when motivational strategies were introduced; however, this effect was not replicated. More specifically, high levels of correct responding did not maintain when the motivational interventions condition was implemented for an extended period of time. Sustained increases in correct responding were observed when the combined phase was implemented, and this effect was replicated following returns to the motivational interventions and prompts conditions.

For Liv (Figure 7), an increase in correct responding was observed during the prompts condition for her block-structure task; however, the effectiveness of the prompts condition was not replicated following a return to the motivational interventions condition. Given this, the combined phase was introduced, and high and sustained levels of correct responding were observed. This effect was replicated following returns to both the motivational interventions and prompts conditions. For her remaining two tasks (screw and shape sorter), a gradual increase in correct responding was observed during the initial combined phase. For the screw task, levels of correct responding remained high when prompts were removed and motivational strategies alone were implemented. However, a higher and more stable level of correct responding was observed when the combined phase was reintroduced. This effect was replicated following a return to the prompts condition. Given that responding maintained at relatively high levels in the second motivational interventions condition, a final motivational interventions condition was conducted. High and stable levels of correct responding were observed during this condition, suggesting that the combination of prompts and motivational strategies were necessary in order to produce an initial increase in correct responding. However, following exposure to the combined condition, high levels maintained when the prompts were removed and motivational strategies alone were implemented. For her shape sorter task, high and stable levels of correct responding were observed during the combined phase relative to the levels of responding observed during the motivational interventions and prompts conditions.

For Zara (Figure 8), increases in correct responding were observed during the prompts condition for her letter-identification task. This effect was replicated following a return to the motivational interventions condition. For Zara's shape sorter and puzzle tasks, however, none of the conditions were sufficient to produce sustained increases in correct responding. Zara's data

suggest that modifications to the motivational or prompting conditions may be necessary for some participants in order for increases in correct responding to be observed. Alternatively, this pattern of responding, with respect to some tasks, may suggest that the participant lacks important prerequisite skills that are necessary in order to demonstrate the desired behavior.

Discussion: Study 1

Results suggested that the procedures developed by Lerman et al. (2004) were effective at evaluating performance on educational tasks and identifying an effective teaching intervention for 20 of the 23 tasks. This study extended the results obtained by Lerman et al. to a younger population of children with no known diagnoses. Similar to the results obtained by Lerman et al., idiosyncratic outcomes were observed across participants and skills. Detailed information regarding the interventions identified as effective for each of the participants can be found in Table 2. For three of the 20 tasks, motivational strategies alone were sufficient to increase correct responding. These results suggest that motivational strategies should be implemented when presenting these tasks in the classroom and that an emphasis should be placed on thinning the use of these strategies (e.g., thinning the reinforcement schedule, decreasing the ratio of known to unknown tasks) while maintaining high levels of correct responding. For six of the 20 tasks, prompting procedures were necessary to increase correct responding. Further instruction with prompts and prompt fading would be recommended for these tasks. For nine of the 20 tasks, a combination of prompting and motivational strategies was necessary to increase correct responding. Thus, teaching in the classroom should involve the use of prompts and motivational strategies and systematic strategies to thin their use. Finally, for two of the 20 tasks, the combined condition was necessary in order to produce an increase in correct responding that approached clinically significant levels. Following exposure to the combined condition,

however, relatively high levels maintained when prompts were removed and motivational strategies alone were implemented. Further instruction with motivational strategies would be recommended with strategies to systematically thin their use.

Although systematic patterns of responding were observed across tasks and participants, our conclusions are limited to the prompting and motivational strategies that were evaluated. High levels of performance may have occurred and maintained in the prompts condition, for example, if alternative prompting procedures had been utilized. Thus, the results of this assessment are limited to the specific teaching strategies that were examined. Additionally, differential lengths of assessment phases may have influenced the results. Extended implementations of each condition may have resulted in either (a) ultimately high levels of performance or (b) poor maintenance of responding. Although extended phases would impact the efficiency of the assessment procedure, they may have been useful in determining the long-term efficacy of each intervention.

While the procedures developed by Lerman et al. (2004) may provide a means to identify performance, skill, or a combination of deficits such that relevant teaching strategies can be recommended, some procedural limitations prohibit making definitive conclusions. First, motivational strategies alone may be sufficient to quickly teach skills. In these situations, skill deficits may be masked as performance deficits. Second, the use of prompting procedures may function in some situations as a motivational strategy by minimizing response effort. Thus, performance deficits may be masked as skill deficits. Additional information could be obtained by counterbalancing the order of the motivational interventions and prompts conditions such that some skills are exposed to prompting prior to motivational strategies. Future researchers also

could expose tasks to each assessment condition regardless of responding to obtain additional information about responding under each of the experimental conditions.

Lerman et al. (2004) emphasized not only the effectiveness of this assessment procedure at evaluating performance on educational tasks but also the ability to do so in an efficient manner. However, roughly 17 to 70 sessions were required per task to demonstrate systematic changes in responding for the participants in Study 1. In comparison to the assessment procedures that utilize a brief or standard multielement design, the number of sessions that were necessary to complete the analysis may make mention of efficiency debatable.

We reanalyzed the data from Study 1, for the 20 tasks for which an effective intervention was identified, to compare the pattern of responding during the initial presentation of each of the experimental conditions. This analysis mimicked the type of graphical depiction that would be obtained by using a brief multielement assessment. Examples of this analysis are depicted in Figures 9 and 10. Beth's responding to the puzzle task using the standard reversal design is depicted in the top panel of Figure 9. These same data are depicted in the bottom panel with only the initial data point included. In this example, the same conclusion could be drawn based on visual inspection of eight data points as compared to that of 39. Ivy's responding to the object-identification task is depicted in the top panel of Figure 10. The bottom panel depicts these same data including only the initial data point from each of the assessment conditions. In this example, the prompts condition would not have been identified as effective, and additional interventions would have been evaluated. Out of the 20 data sets in Study 1 for which an effective intervention was identified, the single-point analysis produced a match in interpretation for nine of those data sets.

However, there are significant limitations to this type of retrospective analysis that limit the conclusiveness of the results. First, repeated exposure to the task and each of the assessment conditions in Study 1 may have differentially influenced responding in subsequent conditions. For example, repeated exposure to baseline sessions may have made the introduction of motivational strategies more salient to the participant, producing a more immediate change in performance. Alternatively, repeated exposure to prompting procedures in the absence of motivational strategies may have been necessary before observing disruption in performance. Extended exposure to prompts in the absence of reinforcement may have enhanced the discriminability of this arrangement when the prompts condition was implemented in future phases. Second, decisions regarding treatment effectiveness were based off visual inspection of the entire data path. In many cases, the same sequence of conditions would not have been conducted based solely on the results of the single-point analysis. Given this, the logic by which phases were conducted was not consistent across tasks and participants, and in many cases, inconclusive results were obtained, but data were not available with respect to responding under subsequent experimental conditions. Given issues with efficiency in Study 1 and the inconclusive results that could be obtained by retrospective analysis, the purpose of Study 2 was to conduct systematic assessments, across a number of target tasks, utilizing three different experimental designs: brief multielement, standard multielement, and standard reversal. Subsequently, data were analyzed to determine the assessment procedure that most efficiently and consistently identified interventions with long-term efficacy.

Method: Study 2

Participants and Setting

Participants were three toddlers with no known diagnoses, ranging in age from 18 to 25 months. All participants attended a university-based daycare classroom. Like Study 1, each of the participants had mastered a number of skills in their classroom curriculum prior to their participation in the study. Detailed information regarding each of the participants can be found in Table 3. Sessions were conducted in the same setting as described in Study 1.

Response Measurement and Reliability

Data were collected on the frequency of target instructions, correct responding, maintenance instructions, prompts, praise, and the delivery of preferred items as defined in Study 1. As in Study 1, responding that occurred after the experimenter prompt was considered correct. A second observer simultaneously, but independently, collected data during at least 48% of the sessions for each participant. Interobserver agreement was calculated in the manner described in Study 1. Mean interobserver agreement for target instructions and correct responding was 99% (range, 16% to 100%) and 99% (range, 25% to 100%), respectively. Mean interobserver agreement for the delivery of maintenance instructions, prompts, praise, and preferred items was 99% (range, 21% to 100%), 99% (range, 26% to 100%), 99% (range, 82% to 100%), and 99% (range, 40% to 100%), respectively. Treatment integrity was calculated for each session as described in Study 1. Treatment fidelity was 100%.

Procedure

Pre-Test. Potential target tasks were selected for each participant from the participant's classroom curriculum or popular curriculum guides (based on the participant's age, classroom observations, and previously mastered skills). Pre-test baseline sessions were conducted as described in Study 1 to identify appropriate tasks for inclusion in Study 2. Tasks were included in Study 2 if the mean level of correct responding in baseline fell at or below 50%. Baseline

sessions continued to be conducted until six target tasks were identified for each participant. Of the six target tasks, two primary skills (e.g., sorting, completing a puzzle) were selected, and three subsets of exemplars were identified. For example, sorting colors may constitute target task 1 while target tasks 2 and 3 may involve sorting shapes and objects, respectively.

Preference Assessment. A 10-item paired-choice preference assessment was conducted, using procedures similar to those described by Fisher et al. (1992), to identify highly preferred edible items for each participant. The highest ranked items were selected for use in the motivational interventions and combined conditions as described in Study 1. As in Study 1, selected items were unavailable to the participant in the classroom outside of experimental sessions, with a few exceptions. First, if the child was simultaneously participating in an intensive toilet-training regimen, some of the same items may have been used as reinforcers in that setting. Second, Cheerios® for Bo were occasionally served during breakfast in the classroom.

Skill Assessment. Skill assessments were conducted with each of the target tasks. Performance was assessed during baseline, motivational interventions, prompts, and combined conditions as described in Study 1. Performance was assessed using three types of experimental designs: a brief multielement, a standard multielement, and a standard reversal.

Two target tasks, addressing different primary skills (e.g., completing block designs by color and placing a star and pentagon in a shape sorter), were randomly selected for assessment using the brief multielement design. A single session was conducted in each condition to identify the least intensive intervention that resulted in correct responding during a minimum of 80% of trials. Subsequently, a single reversal session was conducted using (a) baseline conditions or (b) one of the intervention conditions identified to be ineffective. Following this

session, the potentially effective treatment was reintroduced. If high levels of correct responding ($\geq$ 80%) were not replicated, a replication was attempted with a condition identified as more intensive but similarly effective. The results of the brief multielement assessment were considered inconclusive if replications of treatment effects could not be obtained under any of the treatment conditions or if clinically significant levels of correct responding (i.e., 80% of trials) were not observed during any of the assessment conditions.

Two target tasks, addressing different primary skills (e.g., completing block designs by shape and placing a diamond and triangle in a shape sorter), were randomly selected for assessment using the standard multielement design. The four conditions (baseline, motivational interventions, prompts, combined) were rapidly alternated in a series. The order of conditions within each series was randomly determined. Sessions were conducted until differentiated responding between test and control conditions was observed or undifferentiated responding continued to be observed following a maximum of six series, whichever occurred first. The least intensive condition that produced (a) high levels of correct responding and (b) differentiated responding relative to baseline was identified as the effective intervention. The results of the standard multielement assessment were considered inconclusive if undifferentiated responding continued to be observed across a maximum of six series.

Two target tasks, addressing different primary skills (e.g., completing block designs by size or placing a square and triangle in a shape sorter), were randomly selected for assessment using the standard reversal design. Initially, conditions were introduced sequentially, and repeated measures of responding were obtained during each condition. The least intensive condition that produced high levels of correct responding was identified. At that point, a replication of the observed effect was conducted, similar to the manner described in Study 1.

The least intensive condition that produced consistently higher levels of correct responding relative to the other conditions was considered the effective treatment. Results from the standard reversal assessment were considered inconclusive if high levels of responding failed to maintain under any of the conditions.

In order to minimize the likelihood that a history of exposure to the assessment conditions would differentially influence the clarity of the assessment results, all assessments were conducted concurrently. In other words, the attempt was to avoid a situation in which the clarity of the assessment results was influenced by the sequence in which the assessments were conducted. Alternatively, it would be possible to observe enhanced clarity during one assessment simply because it followed completion of a former assessment, in which a history of exposure was provided to each of the assessment conditions. By conducting the assessments concurrently, a history of exposure to the assessment conditions should have affected the results of each assessment in a similar manner.

For any assessment in which inconclusive outcomes were observed, the stimuli were evaluated again using an assessment design that did produce differentiated outcomes for that target skill. Inconclusive outcomes were only observed on two instances. For Bo's puzzle task, the brief and standard multielement assessments failed to identify an effective intervention. However, systematic patterns of responding were observed during the standard reversal assessment for the puzzle task. Given this, the stimuli assigned to the brief and standard multielement assessments were reevaluated using the standard reversal assessment. This was done to try and determine whether undifferentiated outcomes were a function of the stimuli assessed or a function of the experimental design that was utilized.

Extended Evaluation. For any skill assessment in which clear and differentiated outcomes were observed, an extended evaluation was conducted to further validate the assessment results. The extended evaluation began by conducting repeated baseline sessions. Following baseline, a multielement evaluation was conducted to compare responding during implementation of a matched (i.e., effective) and mismatched (i.e., ineffective) intervention, as identified from the skill assessment. The matched intervention was selected by identifying the least intensive assessment condition that produced the most consistent and clinically significant increase in correct responding. Selection of the mismatched intervention was a bit more discretionary. If the same interpretation could be made across assessments for a given skill (e.g., the combined condition always produced the highest level of responding, regardless of the exemplars assessed or the experimental design that was utilized), we ensured that different assessment conditions (i.e., prompts or motivational interventions) were selected as the mismatched intervention across the subsets of exemplars. In many cases, the intervention that produced the second highest level of correct responding was selected for evaluation. When the combined condition was identified across assessments as being most effective, the motivational interventions condition was often selected as the mismatched intervention. This was done to determine whether a history of exposure to the combined condition would produce an increase in independent responding over time when prompts were removed and motivational strategies alone were implemented. If elevated levels of correct responding were observed during all assessment conditions, we selected the least intensive conditions (i.e., baseline and motivational interventions) for the extended evaluation. If elevated levels of correct responding were observed during all of the assessment conditions excluding baseline, the use of prompts seemed superfluous. Thus, the motivational interventions condition was selected as the matched

intervention, and the baseline condition was selected as the mismatched intervention. For any skill assessment in which high and stable levels of correct responding were observed in the motivational interventions condition relative to the prompts condition, the prompts condition was selected as the mismatched intervention. There was never a situation in which the motivational interventions condition produced optimal responding relative to the combined condition; thus, the prompts condition was the only suitable comparison. After every fourth session in the multielement evaluation, a baseline probe was conducted. This was done to assess periodically the level of correct responding in the absence of all instructional and motivational strategies.

Data Analysis

Data from the skill assessments and extended evaluations were analyzed by comparing the percentage of trials with correct responding across the experimental conditions. Assessment efficiency was examined by calculating the duration (in min) of each assessment. The mean duration to completion for each of the experimental designs was calculated by summing the duration of each assessment (for which an effective intervention was identified), across skills and participants for a given experimental design, and dividing by the total number of assessments utilizing that experimental design. Similar calculations were conducted to determine the mean number of sessions to completion and the mean number of trials to completion across the three assessment designs. The brief multielement assessment required an average of 27 min (range, 18 to 30), 109 trials (range, 87 to 114), or 8 sessions (range, 6 to 8) to complete. The standard multielement assessment required an average of 55 min (range, 34 to 83), 226 trials (range, 134 to 321), or 17 sessions (range, 10 to 25) to complete. The standard reversal assessment required an average of 81 min (55 to 106), 343 trials (215 to 404), or 27 sessions (18 to 31) to complete. Thus, the brief multielement assessment was the most efficient assessment arrangement.

Detailed information regarding efficiency outcomes across participants, tasks, and experimental designs can be found in Table 4.

Assessment efficacy was examined by determining the percentage of cases in which the results of the skill assessment accurately predicted (a) an intervention with long-term effectiveness (i.e., matched intervention) and (b) an intervention that failed to maintain high levels of performance (i.e., mismatched intervention). In order to conclude that correspondence was observed between the results of the skill assessment and those of the extended evaluation, (a) at least one assessment condition must produce consistent and clinically significant levels of correct responding during the skill assessment, (b) differentiated responding between the matched and mismatched interventions, in the direction predicted by the assessment, must be observed during the extended evaluation, and (c) the mean level of correct responding in the matched condition of the extended evaluation must meet or exceed 80% (i.e., clinically significant levels). Using the criteria above, the brief multielement assessment successfully identified effective and ineffective interventions in three of six opportunities. The standard multielement and standard reversal identified effective and ineffective interventions in three of six and six of eight opportunities, respectively. Thus, the standard reversal assessment was identified as being the most efficacious assessment arrangement. Detailed information for each of the participants regarding the correspondence between the results of their skill assessments and the results of their extended evaluations can be found in Table 5.

Results: Study 2

Results for Study 2 are depicted in Figures 11 through 17. Sessions are depicted along the x-axis and the percentage of trials with correct responses along the y-axis. Results of the skill assessments are depicted in the left-hand column. Results from the brief multielement are

depicted in the top panel, standard multielement in the middle panel, and standard reversal in the

bottom panel. The results of the corresponding extended evaluations are depicted in the right-

hand column. For Bo's animal-identification task (Figure 11), a gradual increase in the level of

correct responding was observed across all conditions of the brief multielement assessment (i.e.,

identifying squid, skunk, and snail). These data suggested that the discriminations had been

acquired, and prompting was no longer warranted. Given this, we evaluated responding during

the two conditions (i.e., baseline and motivational interventions) that excluded prompts in the

extended evaluation. High and stable levels of correct responding were observed in the

motivational interventions condition relative to the level of responding observed in the baseline

condition. Thus, although the skill assessment suggested that high levels of performance should

maintain under baseline conditions, results of the extended evaluation indicated that more

systematic thinning of the motivational strategies was warranted. Given this, we concluded that

correspondence was not observed between the results of the skill assessment and the results of

the extended evaluation. During the standard multielement assessment (i.e., identifying

porcupine, peacock, and parrot), high and stable levels of correct responding were observed

during the combined, prompts, and motivational interventions conditions during the last two

series of the assessment. In contrast, variable levels of correct responding were observed during

baseline. During the extended evaluation, differentially higher levels of correct responding were

observed during the motivational interventions condition (i.e., matched intervention) as

compared to baseline (i.e., mismatched intervention); however, the mean level of correct

responding during the motivational interventions condition fell at 77%, slightly below our

percentage indicating clinical significance. Thus, responding failed to meet our criteria for

correspondence. During the standard reversal assessment (i.e., identifying lizard, lobster, and

llama), elevated levels of correct responding were observed during each implementation of the motivational interventions condition and the second implementation of the prompts condition. Variable levels of correct responding were observed during baseline. During the extended evaluation, we compared responding during the motivational interventions condition (i.e., matched intervention) to responding in baseline (i.e., mismatched intervention). Although differentiation between these data paths was observed, the mean level of correct responding during the matched condition fell at 73%, failing to meet the criteria for correspondence. Overall, correspondence criteria were not met across any of the experimental designs for Bo's animal-identification task.

For Bo's puzzle task (Figure 12), the brief multielement assessment (i.e., placing pajamas and book pieces in the puzzle) failed to identify an intervention that produced high and stable levels of correct responding. During the standard multielement assessment (i.e., placing bed and pillow pieces in the puzzle), variable levels of correct responding were observed across all assessment conditions. Thus, an effective intervention could not be identified. During the standard reversal assessment (i.e., placing robe and blanket pieces in the puzzle), high and stable levels of correct responding were observed during each implementation of the combined condition. After observing low levels of correct responding during the initial baseline sessions of the extended evaluation, differentially higher levels of correct responding were observed in the combined condition (i.e., matched intervention) as compared to the level of responding in the prompts condition (i.e., mismatched intervention) and baseline probes. Thus, correspondence was observed between the results of the skill assessment and the results of the extended evaluation. Given clarity of the assessment results using the standard reversal design, the other sets of stimuli (i.e., pajamas and book, bed and pillow) were reevaluated using the standard

reversal design (middle panels of Figure 13). Across both skill assessments, the motivational interventions condition was the least intensive condition that produced clinically significant levels of correct responding. During the extended evaluation, elevated levels of correct responding were observed during the motivational interventions condition (i.e., matched intervention) relative to the level of responding in baseline (i.e., mismatched intervention) across both sets of stimuli (bottom panels of Figure 13). Thus, when analyzed using the standard reversal design, the results for both sets of stimuli met the criteria for correspondence. However, it is unclear whether this enhanced clarity was a function of the experimental design or whether enhanced clarity would have been observed under repeated administration of any of the assessment arrangements as a function of repeated exposure to the task and the assessment conditions.

For Hunter's sorting task (Figure 14), high and stable levels of correct responding were observed during each implementation of the combined condition during the brief multielement assessment (i.e., sorting shapes). Following low levels of correct responding in baseline, high and stable levels of correct responding were also observed during the combined condition (i.e., matched intervention) of the extended evaluation as compared to moderate and variable levels of correct responding in the prompts condition (i.e., mismatched intervention) and baseline probes. During the standard multielement assessment (i.e., sorting colors), the combined condition was the only condition that produced high and stable levels of correct responding. During the extended evaluation, low levels of correct responding were observed during the initial baseline phase. Following baseline, differentially higher levels of correct responding were observed in the combined condition (i.e., matched intervention) relative to the motivational interventions condition (i.e., mismatched intervention). During the standard reversal assessment (i.e., sorting

43

characters), elevated levels of correct responding were observed during each implementation of the combined condition. After observing near-zero levels of correct responding in the baseline sessions of the extended evaluation, differentially higher levels of correct responding were observed during the combined condition (i.e., matched intervention) relative to the motivational interventions condition (i.e., mistmatched intervention) and baseline probes. Overall, correspondence between the results of the skill assessments and the results of the extended evaluations were observed for all three subsets of sorting stimuli across the three experimental designs.

For Hunter's puzzle task (Figure 15), the combined condition was the only condition that produced high and stable levels of correct responding during the brief multielement assessment (i.e., placing cloud, wind, and sun pieces in the puzzle). Near-zero levels of correct responding were observed following a return to baseline during the extended evaluation. Subsequently, the combined condition (i.e., matched intervention) produced high and stable levels of correct responding as compared to moderate levels of correct responding in the motivational interventions condition (i.e., mismatched intervention) and baseline probes. During the standard multielement assessment (i.e., placing snow, rainbow, and tornado pieces in the puzzle), high and sustained levels of correct responding were observed during each implementation of the combined condition. Low levels of correct responding were observed during the baseline sessions of the extended evaluation. Following baseline, perfect performance was observed during the combined condition (i.e., matched intervention) as compared to moderate levels of correct responding in the motivational interventions condition (i.e., mismatched intervention) and variable levels of correct responding in the baseline probes. During the standard reversal assessment (i.e., placing lighting, ice, and rain pieces in the puzzle), stable and clinically

significant levels of correct responding were observed during each implementation of the combined condition. Although elevated levels of correct responding were observed during the prompts and motivational interventions conditions, this effect failed to maintain at clinically significant levels over repeated observations. During the extended evaluation, moderate levels of correct responding were observed during baseline. Following baseline, high levels of correct responding were observed in the combined condition (i.e., matched intervention) relative to moderate levels of correct responding in the prompts condition (i.e., mismatched intervention) and baseline probes. Thus, correspondence was observed between the results of the skill assessments and the results of the extended evaluations for all three subsets of puzzle stimuli across the three experimental designs.

For Xander's block-design task (Figure 16), high and stable levels of correct responding were observed during the combined condition of the brief multielement assessment (i.e., block designs by color). Near-zero levels of correct responding were observed during the baseline sessions of the extended evaluation. Subsequently, elevated levels of correct responding were observed during the combined condition (i.e., matched intervention) as compared to the level of responding in the prompts condition (i.e., mismatched intervention) and baseline probes. Thus, correspondence was observed across the skill assessment and extended evaluation using the brief multielement assessment. During the standard multielement assessment (i.e., block designs by shape), stable and clinically significant levels of correct responding were observed during each implementation of the combined condition. Low to moderate levels of correct responding were observed during the baseline sessions of the extended evaluation. Following baseline, high and stable levels were observed in the combined condition (i.e., matched intervention) as compared to the level of responding observed in the motivational interventions condition (i.e., mismatched

45

intervention) and baseline probes. Thus, correspondence was also observed across the skill assessment and extended evaluation using the standard multielement assessment. During the standard reversal assessment (i.e., block designs by size), consistent increases in correct responding were observed during each implementation of the combined condition. After observing low levels of correct responding in baseline, differentially higher levels of correct responding were observed in the combined condition (i.e., matched intervention) relative to the motivational interventions condition (i.e., mismatched intervention) and baseline probes during the extended evaluation. However, the mean level of correct responding in the matched condition of the extended evaluation was 79%, failing to meet the criteria for correspondence.

For Xander's shape sorter task (Figure 17), high and stable levels of correct responding were observed during each implementation of the combined condition during the brief multielement assessment (i.e., inserting star and pentagon pieces). During the extended evaluation, moderate levels of correct responding were observed during the initial baseline phase. Subsequently, high and stable levels of correct responding were observed during both the combined (i.e., matched intervention) and motivational interventions (i.e., mismatched intervention) conditions. Given that the mismatched intervention failed to produce reduced levels of correct responding, correspondence was not observed across the skill assessment and extended evaluation. During the standard multielement assessment (i.e., inserting diamond and triangle pieces), high and stable levels of correct responding were observed during the combined condition. An increase in the level of correct responding was observed during the motivational interventions condition over the course of sessions. In contrast, moderate and variable levels of correct responding were observed during the prompts and baseline sessions. During the extended evaluation, low levels of correct responding were observed during the initial baseline

phase. Following baseline, low to moderate levels of correct responding were observed during the initial sessions in which motivational interventions (i.e., matched intervention) were implemented. An increase in the level of correct responding was observed in this phase following the first two sessions. In contrast, moderate levels of correct responding were observed during the prompts condition (i.e., mismatched intervention) and baseline probes. Given that the matched condition produced a delayed increase in correct responding and that the mean level of correct responding fell at 68%, correspondence between the results of the skill assessment and the results of the extended evaluation were not observed. During the standard reversal assessment (i.e., inserting square and triangle pieces), elevated levels of correct responding were observed during each implementation of the motivational interventions condition as compared to the level of responding in the prompts and baseline conditions. Moderate levels of correct responding were observed during the baseline sessions of the extended evaluation. Following baseline, high and stable levels of correct responding were observed in the motivational interventions condition (i.e., matched intervention) as compared to the level of responding observed in the prompts condition (i.e., mismatched intervention) and baseline probes. Thus, correspondence was observed across the skill assessment and extended evaluation.

Discussion: Study 2

Results of Study 2 suggested that developing assessment conditions to test broadly for potential deficits was effective at producing systematic responding during 18 of 20 skill assessments. These outcomes were observed across participants, tasks, and experimental designs. Correspondence between these assessment results and the results of the extended evaluations were observed during 12 of 18 opportunities. The brief and standard multielement

assessments each produced correspondence with the extended evaluation in three of five opportunities. The standard reversal produced correspondence with the extended evaluation in six of eight opportunities, suggesting that the standard reversal was the best predictor of treatment efficacy.

Several patterns of responding were common when correspondence was not observed. First, for three of the 12 instances in which correspondence did not occur (i.e., Xander's completion of block designs by size; Bo's identification of porcupine, peacock, and parrot; and Bo's identification of lizard, lobster, and llama), differentiation was observed between the matched and mismatched interventions in the direction predicted by the assessment, but the mean level of correct responding in the matched intervention fell below 80%. Given that the matched intervention for each of these evaluations involved the delivery of edible items contingent on correct responding, these data may suggest an issue with reinforcer satiation. Given that the extended evaluations occurred after the skill assessments, each of the participants had experienced a long exposure to the reinforcement contingency. After participating in more than 100 experimental sessions, the potency of the preferred items as reinforcers for these tasks may have decreased. It may have been worthwhile to experimentally examine this possibility by introducing a new array of preferred items during the extended evaluation and observing its effect on responding. However, variation in the preferred items also may have affected the results of the skill assessment; thus, this examination was not conducted. Future researchers may be interested in conducting more frequent preference assessments to minimize the potential for reinforcer satiation, especially when a large number of sessions are expected from the onset of the study.

Second, for Bo's identification of squid, snail, and skunk and Xander's placement of a star and pentagon in a shape sorter, the brief multielement assessments may have involved too few exposures to the assessment conditions to identify stable patterns of responding. In Xander's case, the brief multielement assessment failed to identify the motivational interventions condition as being effective. This result was not entirely surprising given the pattern of responding observed during his skill assessment. During the first exposure to the motivational interventions condition, Xander responded correctly on 10% of opportunities. During his second exposure to this condition, Xander's level of correct responding increased to 60%. In situations like this, where a sizeable increase in the level of correct responding is observed across implementations of the same condition, additional implementations likely are warranted to determine whether further increases in correct responding will be observed. In Xander's case, careful visual inspection would likely have encouraged further analysis. This is a case in which visual inspection should have guided assessment length rather than using a predetermined criterion of clinical significance. In Bo's case, we observed a gradual increase in the level of correct responding over the course of sessions, ending with 100% correct responding during baseline. These data suggested that the discriminations had been acquired, and prompting was no longer warranted. However, this assessment likely involved too few exposures to baseline conditions to determine conclusively whether this newly acquired response would maintain at high levels in the absence of motivational strategies. Thus, the assessment may have contained insufficient information to identify the baseline condition as a matched intervention. In response to similar patterns of responding in future skill assessments, the recommendation likely would be to implement the motivational strategies in the classroom and systematically fade their use.

Xander's placement of a diamond and triangle in a shape sorter was the last situation in which correspondence was not observed between the skill assessment and extended evaluation. During the skill assessment, we observed a gradual increase in the level of correct responding during the motivational interventions condition. Within-session data suggested that errors and non-responding were gradually eliminated, and more fluent responding likely was shaped over the course of sessions. Following the skill assessment, three baseline sessions were implemented in which motivational strategies were removed. Within-session data suggested that correct responding during these sessions was replaced with attempts to respond correctly (i.e., Xander was making attempts to perform the task correctly but was not completing the task before the trial elapsed). Thus, the fluent responding that had been established during the skill assessment did not maintain. This information may suggest why an immediate change in performance was not observed when the motivational interventions condition was reintroduced during the extended evaluation. Within-session data from the multielement component of the extended evaluation suggested that several sessions of the matched intervention were required before fluent responding was reestablished. Despite failure to immediately replicate the level of responding observed at the end of the skill assessment, potential issues with disrupted fluency are unlikely to be encountered under more typical teaching arrangements. In our study, the series of baseline sessions was included in the extended evaluation to enhance the experimental rigor of the evaluation. We were interested in examining the level of correct responding in the absence of instructional and motivational strategies following a history of exposure to these teaching interventions. Under more typical assessment arrangements (i.e., assessments conducted by teachers or clinicians), these baseline sessions would likely be omitted, and the

effective intervention would immediately be introduced in the classroom, minimizing issues with disrupted fluency.

In addition to examining the validity of the assessment results, Study 2 examined the efficiency of the assessment procedures. Not surprisingly, the brief multielement assessment was identified as the most efficient arrangement across all measures (i.e., duration in min to assessment completion, number of trials to assessment completion, number of sessions to assessment completion). Somewhat surprisingly, the standard multielement assessment required more time to complete than the brief multielement assessment but produced no greater correspondence between the results of the skill assessment and those of the extended evaluation. Taken together, this information seems to suggest that modifications be examined to the brief multielement assessment to enhance its efficacy while maintaining its efficiency.

Based on our data from Study 2, modifications to the brief multielement assessment seem both reasonable and promising. There were three situations in which the brief multielement assessment either (a) failed to identify an effective intervention (i.e., Bo's puzzle task) or (b) lacked correspondence with the results of the extended evaluation (i.e., Bo's animal-identification task and Xander's shape sorter task). In each situation, minor modifications could have been made to enhance the accuracy of the assessment results. First, when the brief multielement assessment was used to evaluate Bo's completion of a puzzle, none of the assessment conditions produced high and sustained levels of correct responding. The most logical and time efficient manipulation would have been to conduct an extended series of combined sessions rather than conducting the assessment anew using the standard reversal design. If responding failed to reach high levels during repeated implementation of the combined condition, modifications to the prompting and reinforcement procedures should be

evaluated. If low levels of responding continued to be observed, these data likely would suggest that the student lacked important prerequisite skills that were necessary to complete the task correctly. Alternatively, if correct responding increased and maintained during the extended combined phase, the combined intervention could be implemented in the classroom with strategies to fade the instructional and motivational components. Second, when the brief multielement assessment was used to examine Xander's responding to the shape sorter task, more careful inspection of the data should have been conducted to determine an appropriate assessment length. As mentioned previously, an additional exposure to the motivational interventions condition may have been sufficient to identify it as the most effective and least intensive teaching strategy. Additionally, given the rapidity with which independent responding emerged in the motivational interventions condition, these data highlight the importance of systematically fading the intervention components as soon as the treatment is implemented in the classroom. Finally, when the brief multielement assessment was used to examine Bo's performance during the animal identification task, a gradual increase in correct responding was observed over the course of sessions. Rather than selecting the baseline arrangement as the most effective and least intensive intervention, the motivational interventions condition should have been selected for implementation in the classroom with procedures for systematically fading the motivational components.

Although the results from Study 2 identified one assessment design as being most efficient and another as being most efficacious, the obtained results were helpful in identifying modifications that could be made to enhance the accuracy of the more efficient arrangement. In other words, the results from Study 2 helped identify experimental progressions that could be implemented, given various patterns of responding, in order to enhance the utility of the brief

multielement assessment. Such information should enhance the practicality of conducting these experimental analyses to identify effective interventions for improving academic performance.

General Discussion

This series of studies attempted to address several questions. Results of Study 1 replicated the usefulness of the procedures developed by Lerman et al. (2004) at clearly evaluating the pre-academic performance of toddlers with no known diagnoses. In contrast to the results obtained by Lerman et al. in which the combined condition was required in only 25% of evaluations, the combined condition was necessary, at least initially, in 55% of the evaluations conducted in Study 1. Given the young age of our participants (16 to 26 months), they may have (a) presented with more limited skill repertoires or (b) had a shorter history with direct instruction and discrete-trial training than the participants in their study. Given our results, it may be argued that selecting the combined condition for all skills would be more practical and efficient than conducting this type of systematic assessment across skills and participants. However, the major limitation of this approach is that skills likely will be targeted for intensive teaching that require only the implementation of motivational strategies. Thus, this approach would minimize the number of appropriately challenging instructional goals that could be targeted simultaneously, delaying acquisition of more advanced skills. With this approach, it would be necessary to provide exposure to the combined condition and begin fading the instructional and motivational components before identifying high levels of performance in the absence of prompts. If a more efficient assessment design were utilized, the time required to provide this exposure would likely meet or exceed the amount of time required to complete the assessment.

The time intensive nature of the assessment procedure used in Study 1 was of large concern. In Study 1, between 17 and 70 sessions were required per task in order to demonstrate systematic patterns of responding. This was particularly troublesome given that the goal of this assessment was to identify effective teaching interventions. After a large number of sessions, the experimenters had simply identified a teaching strategy that would be effective in producing the desired response. Fading of these strategies was still warranted. Thus, the time intensive nature of this assessment procedure is particularly problematic because it limits the practicality of conducting systematic skill assessments across a large number of students and educational tasks.

Although Lerman et al. (2004) effectively extended the skill assessment methodology to a novel dependent variable and a novel population of participants, they utilized a more thorough and time-intensive experimental arrangement than was common in previous literature. This assessment arrangement provided a more conservative evaluation of the effectiveness of their assessment approach but may have limited the widespread adoptability of their procedures. Given the clear and promising results obtained in Study 1, Study 2 was designed to address the issue with efficiency to try and enhance the adoptability of the assessment procedure. Study 2 compared the efficiency and validity of a brief multielement assessment, standard multielement assessment, and standard reversal assessment. Results of Study 2 suggested that the brief multielement was the most efficient arrangement in terms of duration to assessment completion, number of sessions to assessment completion, and number of trials to assessment completion. The standard reversal was identified as being most efficacious, producing the highest percentage of correspondence between the results of the skill assessments and the results of the extended evaluations. Results of Study 2 did not identify a single assessment arrangement that efficiently and consistently identified interventions with long-term efficacy. However, post-hoc evaluations

of situations in which correspondence was not observed suggest that minor modifications could be made to the brief multielement assessment to enhance its predictive validity while maintaining high levels of efficiency. Additional research is warranted, however, to verify the effectiveness of these modifications at enhancing the efficacy of the brief multielement assessment.

An interesting finding from Study 2 was that similar patterns of responding were commonly observed across experimental designs for each of the primary skills. That is, the same intervention was identified as effective across the majority of evaluations, regardless of whether performance for each subset of exemplars was evaluated using a brief multielement, standard multielement, or standard reversal. These results suggest that, for any given participant, there may be categories of tasks that a given intervention is well suited to address. We evaluated similar subsets of exemplars across experimental designs in an effort to determine whether differentiated or undifferentiated assessment outcomes were a function of the design as opposed to differences in the tasks being evaluated. However, it is unclear how the clarity of our assessment results was influenced by this arrangement. We may have observed enhanced clarity across assessments as a function of conducting concurrent sessions with similar task materials that provided (a) repeated practice with a related task or (b) repeated exposure to each of the assessment conditions in the context of a similar skill. Future researchers may be interested in evaluating disparate tasks across experimental designs to minimize the potential for generalized responding, to evaluate the impact on assessment clarity, and to enhance the breadth of skill domains to which the assessment methodology is applied.

Another interesting finding across Study 1 and Study 2 was that clear reversals were observed when the combined condition was identified as being effective. That is, prompting

procedures or motivational strategies were rarely effective at producing high levels of performance when implemented in isolation despite a history of exposure to the combined condition. These data suggest that a history of exposure to both instructional and motivational strategies is rarely sufficient to produce independent responding in the absence of systematic fading procedures. Thus, systematic fading of the instructional and motivational interventions is warranted and should be targeted upon implementation of the effective strategy in the classroom setting.

Although Study 2 provided some interesting preliminary information regarding modifications that could be made to enhance the efficiency and validity of the assessment arrangement, future research will benefit from further extensions of this methodology to other skill domains, particularly the assessment of vocal verbal behavior. Additionally, replications of these results are warranted with other participants. Given the vast amount of previous research that has targeted elementary-aged children with no known diagnoses, additional replications are warranted with (a) younger populations of participants and (b) children with intellectual and developmental disabilities. Future research also will benefit from conducting extended validation procedures under more typical classroom arrangements. In the current study, the experimenter conducted the extended evaluation under the same conditions that were utilized during the skill assessment. This approach may have limited the generality of the findings. By conducting validation sessions under more typical classroom arrangements (i.e., in the classroom environment with classroom staff as implementers), the utility of the assessment procedure may be enhanced.

Finally, future researchers may benefit from identifying dependent variables that can be assessed in small- or large-group formats while continuing to evaluate performance individually.

This information may be of great benefit to teachers and clinicians who lack the time and resources to assess a wide range of skills for a large number of children. The ability to assess performance for a number of individuals simultaneously may further enhance the efficiency and adoptability of this assessment methodology.

References

Ayllon, T., & Kelly, K. (1972). Effects of reinforcement on standardized test performance. *Journal of Applied Behavior Analysis, 5,* 477-484.

Ayllon, T., & Roberts, M. D. (1974). Eliminating discipline problems by strengthening academic performance. *Journal of Applied Behavior Analysis, 7,* 71-76.

Baer, D. M. (2005). Letters to a lawyer. In W. Heward et al. (Eds.), *Focus on behavior analysis in education: Achievements, challenges, and opportunities* (pp. 3-30). Upper Saddle River: Merrill/Prentice Hall.

Bonfiglio, C. M., Daly, E. J., III, Martens, B. K., Lin, L. R., & Corsaut, S. (2004). An experimental analysis of reading interventions: Generalization across instructional strategies, time, and passages. *Journal of Applied Behavior Analysis, 37,* 111-114.

Burns, M. K., Ganuza, Z. M., & London, R. M. (2009). Brief experimental analysis of written letter formation: Single-case demonstration. *Journal of Behavioral Education, 18,* 20-34.

Carr, J. E., Coriaty, S., & Dozier, C. L. (2000). Current issues in the function-based treatment of aberrant behavior in individuals with developmental disabilities. In J. Austin & J. E. Carr (Eds.), *Handbook of applied behavior analysis* (pp. 91-112). Reno, NV: Context Press.

Carr, J. E., Coriaty, S., Wilder, D. A., Gaunt, B. T., Dozier, C. L., Britton, L. N., . . . Reed, C. L. (2000). A review of "noncontingent" reinforcement as treatment for the aberrant behavior of individuals with developmental disabilities. *Research in Developmental Disabilities, 21,* 377-391.

Cuvo, A. J., Ashley, K. M., Marso, K. J., Zhang, B. L., & Fry, T. A. (1995). Effect of response practice variables on learning spelling and sight word vocabulary. *Journal of Applied Behavior Analysis, 28,* 155-173.

Cuvo, A. J., Davis, P. K., O'Reilly, M. F., Mooney, B. M., & Crowley, R. (1992). Promoting

    stimulus control with textual prompts and performance feedback for persons with mild

    disabilities. *Journal of Applied Behavior Analysis, 25,* 477-489.

Daly, E. J., III, Bonfiglio, C. M., Mattson, T., Persampieri, M., & Foreman-Yates, K. (2005).

    Refining the experimental analysis of academic skill deficits: Part 1. An investigation of

    variables that affect generalized oral reading performance. *Journal of Applied Behavior*

    Analysis, 38, 485-497.

Daly, E. J., III, Martens, B. K., Dool, E. J., & Hintze, J. M. (1998). Using brief functional

    analysis to select interventions for oral reading. *Journal of Behavioral Education, 8,* 203-

    218.

Daly, E. J., III, Martens, B. K., Hamler, K. R., Dool, E. J., & Eckert, T. L. (1999). A brief

    experimental analysis for identifying instructional components needed to improve oral

    reading fluency. *Journal of Applied Behavior Analysis, 32,* 83-94.

Daly, E. J., III, Murdoch, A., Lillenstein, L., Webber, L., & Lentz, F. (2002). An examination of

    methods for testing treatments: Conducting brief experimental analyses of the effects of

    instructional components on oral reading fluency. *Education and Treatment of Children,*

    25, 288-316.

Daly, E. J., III, Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a

    functional analysis of academic performance problems. *School Psychology Review,*

    26, 554-574.

Drabman, R. S., & Lahey, B. B. (1974). Feedback in classroom behavior modification: Effects

    on the target and her classmates. *Journal of Applied Behavior Analysis, 7,* 591-598.

Duhon, G. J., Noell, G. H., Witt, J. C., Freeland, J. T., Dufrene, B. A., & Gilbertson, D. N.

(2004). Identifying academic skill and performance deficits: The experimental analysis of brief assessments of academic skills. *School Psychology Review, 33,* 429-443.

Dunlap, G., dePerczel, M., Clarke, S., Wilson, D., Wright, S., White, R., & Gomez, A. (1994). Choice making to promote adaptive behavior for students with emotional and behavioral challenges. *Journal of Applied Behavior Analysis, 27,* 505-518.

Dunlap, G., & Koegel, R. L. (1980). Motivating autistic children through stimulus variation. *Journal of Applied Behavior Analysis, 13,* 619-627.

Eckert, T. L., Ardoin, S. P., Daisey, D. M., & Scarola, M. D. (2000). Empirically evaluating the effectiveness of reading interventions: The use of brief experimental analysis and single case designs. *Psychology in the Schools, 37,* 463-473.

Eckert, T. L., Ardoin, S. P., Daly, E. J., & Martens, B. K. (2002). Improving oral reading fluency: A brief experimental analysis of combining an antecedent intervention with consequences. *Journal of Applied Behavior Analysis, 35,* 271-281.

Elliott, S. N., & Shapiro, E. S. (1990). Intervention techniques and programs for academic performance problems. In T. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (2nd ed., pp. 635-660). New York, NY: Wiley.

Fenerty, K. A., & Tiger, J. H. (2010). Determining preschoolers' preference for choice-making opportunities: Choice of task versus choice of consequence. *Journal of Applied Behavior Analysis, 43,* 503-507.

Fisher, W., Piazza, C. C., Bowman, L. G., Hagopian, L. P., Owens, J. C., & Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis, 25,* 491-498.

Fisher, W. W., Thompson, R. H., Piazza, C. C., Crosland, K., & Gotjen, D. (1997). On the

relative reinforcing effects of choice and differential consequences. *Journal of Applied Behavior Analysis, 30,* 423-438.

Fueyo, V., & Bushell, D. (1998). Using number line procedures and peer tutoring to improve the mathematics computation of low-performing first graders. *Journal of Applied Behavior Analysis, 31,* 417-430.

Gortmaker, V. J., Daly, E. J., III, McCurdy, M., Persampieri, M. J., & Hergenrader, M. (2007). Improving reading outcomes for children with learning disabilities: Using brief experimental analysis to develop parent-tutoring interventions. *Journal of Applied Behavior Analysis, 40,* 203-221.

Hagopian, L. P., Dozier, C. L., Rooker, G. W., & Jones, B. A. (2012). Assessment and treatment of severe problem behavior. In G. J. Madden, W. V. Dube, G. P. Hanley, & K. A. Lattal (Eds.), *APA Handbook of Behavior Analysis* (Vol. 2, pp. 353-386). Washington, D.C.: American Psychological Association.

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

Iwata, B. A., & Dozier, C. L. (2008). Clinical application of functional analysis methodology. *Behavior Analysis in Practice, 1,* 3-9.

Jones, K. M., & Wickstrom, K. F. (2002). Done in sixty seconds: Further analysis of the brief assessment model for academic problems. *School Psychology Review, 31,* 554-568.

Kern-Dunlap, L., Dunlap, G., Clarke, S., Childs, K. E., White, R. L., & Stewart, P. T. (1992).

Effects of a videotape package on the peer interactions of children with serious

behavioral and emotional challenges. *Journal of Applied Behavior Analysis, 25,* 355-

364.

Koegel, L. K., & Koegel, R. L. (1986). The effects of interspersed maintenance tasks on

academic performance in a severe childhood stroke victim. *Journal of Applied*

Behavior Analysis, 19, 425-430.

Kratochwill, T. R., & Sheridan, S. M. (1990). Advances in behavioral assessment. In T. B.

Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (2nd ed., pp. 328-

364). New York: Wiley & Sons.

Lahey, B. B., McNees, M. P., & Brown, C. C. (1973). Modification of deficits in reading for

comprehension. *Journal of Applied Behavior Analysis, 6,* 475-480.

Lerman, D. C., Vorndran, C. M., Addison, L. R., & Kuhn, S. A. C. (2004). A rapid assessment

of skills in young children with autism. *Journal of Applied Behavior Analysis, 37,* 11-26.

Mace, F. C., & Belfiore, P. (1990). Behavioral momentum in the treatment of escape-motivated

stereotypy. *Journal of Applied Behavior Analysis, 23,* 507-514.

Martens, B. K., Eckert, T. L., Bradley, T. A., & Ardoin, S. P. (1999). Identifying effective

treatments from a brief experimental analysis: Using single-case design elements to aid

decision making. *School Psychology Quarterly, 14,* 163-181.

McComas, J. J., Wacker, D. P., Cooper, L. J., Asmus, J. M., Richman, D., & Stoner, B. (1996).

Brief experimental analysis of stimulus prompts for accurate responding on academic

tasks in an outpatient clinic. *Journal of Applied Behavior Analysis, 29,* 397-401.

Neef, N. A., Iwata, B. A., & Page, T. J. (1980). The effects of interspersal training versus high-

density reinforcement on spelling acquisition and retention. *Journal of Applied Behavior Analysis, 13,* 152-158.

Newcomer, L. L., & Lewis, T. J. (2004). Functional behavioral assessment: An investigation of assessment reliability and effectiveness of function-based intervention. *Journal of Emotional and Behavioral Disorders, 12,* 168-181.

Noell, G. H., Ardoin, S. P., & Gansle, K. A. (2009). Academic assessment. In J. L. Matson, F. Adrasik, & M. L. Matson (Eds.), *Assessing childhood psychopathology and developmental disabilities* (pp. 311-340). New York, NY: Springer.

Noell, G. H., Freeland, J. T., Witt, J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students. *Journal of School Psychology, 39,* 335-355.

Noell, G. H., Gansle, K. A., Witt, J. C., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. H., . . . Northup, J. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. *Journal of Applied Behavior Analysis, 31,* 659-663.

Noell, G. H., Roane, H. S., VanDerHeyden, A. M., Whitmarsh, E. L., & Gatti, S. L. (2000). Programming for the generalization of communication to the classroom following assessment and training outside of the classroom. *School Psychology Review, 29,* 429-442.

Odom, S. L., Chandler, L. K., Ostrosky, M., McConnell, S. R., & Reaney, S. (1992). Fading teacher prompts from peer-initiation interventions or young children with disabilities. *Journal of Applied Behavior Analysis, 25,* 307-317.

Partington, J. W. (2006). *The assessment of basic language and learning skills – revised.*

Pleasant Hill, CA: Behavior Analysts, Inc.

Richman, D. M., Wacker, D. P., Cooper-Brown, L. J., Kayser, K., Crosland, K., Stephens, T. J.,
Asmus, J. (2001). Stimulus characteristics within directives: Effects on accuracy of task
completion. *Journal of Applied Behavior Analysis, 34,* 289-312.

Rodgers, T. A., & Iwata, B. A. (1991). An analysis of error-correction procedures during
discrimination training. *Journal of Applied Behavior Analysis, 24,* 775-781.

Roll, D. L. (1973). Modification of nasal resonance in cleft-palate children by informative
feedback. *Journal of Applied Behavior Analysis, 6,* 397-403.

Romaniuk, C., Miltenberger, R., Conyers, C., Jenner, N., Jurgens, M., & Ringenberg, C. (2002).
The influence of activity choice on problem behaviors maintained by escape versus
attention. *Journal of Applied Behavior Analysis, 35,* 349-362.

Schmidt, A. C., Hanley, G. P., & Layer, S. A. (2009). A further analysis of the value of choice:
Controlling for illusory discriminative stimuli and evaluating the effects of less preferred
items. *Journal of Applied Behavior Analysis, 42,* 711-716.

Schulman, J. L., Suran, B. G., Stevens, T. M., & Kupst, M. J. (1979). Instructions, feedback, and
reinforcement in reducing activity levels in the classroom. *Journal of Applied Behavior
Analysis, 12,* 441-447.

Shapiro, E. S., & Derr, T. F. (1990). Curriculum-based assessment. In T. R. Gutkin & C.
Reynolds (Eds.), *The handbook of school psychology* (2nd ed., pp. 365-387). New York,
NY: Wiley.

Sundberg, M. L. (2008). *VB-MAPP: Verbal behavior milestones assessment and placement
program.* Concord, CA: AVB Press.

Tasky, K. K., Rudrud, E. H., Schulze, K. A., & Rapp, J. T. (2008). Using choice to increase on-

task behavior in individuals with traumatic brain injury. *Journal of Applied Behavior Analysis, 41,* 261-265.

Tiger, J. H., Hanley, G. H., & Bruzek, J. (2008). Functional communication training: A review and practical guide. *Behavior Analysis in Practice, 1,* 16-23.

Volkert, V. M., Lerman, D. C., Trosclair, N., Addison, L., & Kodak, T. (2008). An exploratory analysis of task-interspersal procedures while teaching object labels to children with autism. *Journal of Applied Behavior Analysis, 41,* 335-350.

Vollmer, T. R., & Iwata, B. A. (1992). Differential reinforcement as treatment for behavior disorders: Procedural and functional variations. *Research in Developmental Disabilities, 13,* 393-417.

Vollmer, T. R., Iwata, B. A., Duncan, B. A., & Lerman, D. C. (1993). Extensions of multielement functional analysis using reversal-type designs. *Journal of Developmental and Physical Disabilities, 5,* 311-325.

Vollmer, T. R., & Northup, J. (1996). Some implications of functional analysis for school psychology. *School Psychology Quarterly, 11,* 76-92.

Wagner, D., McComas, J. J., Bollman, K., & Holton, E. (2006). The use of functional reading analysis to identify effective reading interventions. *Assessment for Effective Intervention, 32,* 40-49.

Worsdell, A. S., Iwata, B. A., Dozier, C. L., Johnson, A. D., Neidert, P. L., & Thomason, J. L. (2005). Analysis of response repetition as an error-correction strategy during sight-word reading. *Journal of Applied Behavior Analysis, 38,* 511-527.

Participant	Age	Mastered Curriculum Goals
Beth	21 months	Engage in functional play, stack objects, build simple structures, string beads, make marks on paper, vocalize repetitive consonant-vowel combinations, use 10 or more spoken words to label objects or make requests, request assistance by saying "help," make requests using two-word combinations or three-word sentences, recruit teacher attention by saying "hi," take turns with a teacher, take 10 consecutive steps, use a step stool, use a slide, take 10 independent bites, wash hands without assistance, name one color (pink)
Brandy	17 months	Engage in functional play, stack objects, make marks on paper, vocalize repetitive consonant-vowel combinations, use 10 or more spoken words to label objects or make requests, request assistance by saying "help," use plural form of five words, recruit teacher attention by saying "hi," take turns with a teacher, take 10 consecutive steps, use a step stool, use a slide
Eric	21 months	Engage in functional play, stack three objects, make four marks on paper, vocalize repetitive consonant-vowel combinations, use 10 or more spoken words to label objects or make requests, recruit teacher attention by saying "hi," take turns with teachers or peers, take 10 consecutive steps, use a step stool, use a slide, stop and start a gross motor activity when instructed
Ivy	16 months	Engage in functional play, stack objects, string beads, make marks on paper, vocalize repetitive consonant-vowel combinations, use at least four words to label objects or make requests, recruit teacher attention by saying "hi," take turns with a teacher, take 10 consecutive steps, use a step stool, use a slide
Jocelyn	23 months	Engage in functional play, stack objects, build simple structures, string beads, make marks on paper, imitate a 3-step sequence of activities, vocalize repetitive consonant-vowel combinations, use 10 or more spoken words to label objects or make requests, request assistance by saying "help," use two-word combinations or three-word sentences to make requests, recruit teacher attention by saying "hi," take 10 consecutive steps, use a step stool, use a slide, start and stop a gross motor activity when instructed, take 10 independent bites, wash hands without assistance, drink from a cup without a lid, name four colors (blue, yellow, green, pink)
Leanne	19 months	Engage in functional play, stack objects, build simple structures, make marks on paper, vocalize repetitive consonant-vowel combinations, use at least seven spoken words to label objects or make requests, request assistance by saying "help," recruit teacher attention by saying "hi," take turns with teachers and peers, take 10 consecutive steps, use a step stool, use a slide, take 10 independent bites
Liv	26 months	Engage in functional play, stack objects, build simple structures, string beads, make marks on paper, imitate horizontal and vertical strokes, imitate a three-step sequence of activities, vocalize repetitive consonant-vowel combinations, use 10 or more spoken words to label objects or make requests, request assistance by saying "help," use two-word combinations or three-word sentences to make requests, describe own behavior using complete sentence, use plural form of five words, use personal pronouns, recruit teacher attention by saying "hi," take turns with a teacher, take 10 consecutive steps, use a step stool, use a slide, stop and start a gross motor activity when instructed, roll a ball two feet, kick a ball two feet, take 10 independent bites, wash hands without assistance, drink from a cup without a lid, identify 10 colors (pink, red, orange, yellow, green, blue, purple, black, white, brown) and 4 shapes (circle, square, rectangle, triangle), point to name when presented in an array of three
Zara	22 months	Engage in functional play, stack objects, build simple structures, string beads, make four marks on paper, imitate a three-step sequence of activities, vocalize repetitive consonant-vowel combinations, take 10 consecutive steps, use a step stool, take 10 independent bites

Table 1. Names, ages, and mastered curriculum goals for each of the participants in Study 1 at the onset of experimental sessions.

Participant	Intervention Identified as Effective in Skill Assessment				
	MI	Pr	Co	Co → Initial Increase MI → Maintenance	None or Unclear
Beth	1	1	1		
Brandy		2			
Eric	1		1	1	
Ivy		1	2		
Jocelyn	1	1	1		
Leanne			2		1
Liv			2	1	
Zara		1			2
Total	3/23	6/23	9/23	2/23	3/23

Table 2. Teaching interventions identified as effective for each of the participants in Study 1. MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.

Participant	Age	Mastered Curriculum Goals
Bo	18 months	Engage in functional play, stack objects, string beads, make marks on paper, imitate a three-step sequence of activities, request assistance by saying or signing "help," recruit teacher attention by saying "hi," take turns with teachers and peers, take 10 consecutive steps, kick a ball 2 ft, take 10 independent bites
Hunter	25 months	Engage in functional play, stack objects, build simple structures, string beads, make marks on paper, put pieces in a puzzle, vocalize repetitive consonant-vowel combinations, use 10 or more spoken words to label objects or make requests, request assistance by saying "help," use two-word combinations to make requests, recruit teacher attention by saying "hi," take turns with teachers and peers, take 10 consecutive steps, use a step stool, use a slide, roll a ball 2 ft, kick a ball 2 ft, take 10 independent bites, drink from a cup without a lid, match the color green
Xander	23 months	Engage in functional play, stack objects, build simple structures, imitate a three-step sequence of activities, string beads, make marks on paper, put pieces in a puzzle, vocalize repetitive consonant-vowel combinations, use 10 or more spoken words to label objects or make requests, request assistance by saying "help," use two-word combinations to make requests, recruit teacher attention by saying "hi," make requests using "please," say "thank you" after receiving something from others, take turns with teachers and peers, help others without prompting, participate in reciprocal social play, engage in sustained social play, take 10 consecutive steps, use a step stool, use a slide, stop and start a gross motor activity when instructed, roll a ball 2 ft, kick a ball 2 ft, take 10 independent bites, put on a coat, drink from a cup without a lid, name 10 colors, repeat four-component patterns

Table 3. Names, ages, and mastered curriculum goals for each of the participants in Study 2 at the onset of experimental sessions.

Experimental Design	Measures of Assessment Efficiency		
	Duration of Assessment (in min)	Number of Trials to Assessment Completion	Number of Sessions to Assessment Completion
Brief Multielement			
Receptive ID Animals (Bo)	18 min	87 trials	6 sessions
Puzzle (Bo)			
Sort (Hunter)	29 min	114 trials	8 sessions
Puzzle (Hunter)	28 min	114 trials	8 sessions
Block Design (Xander)	29 min	114 trials	8 sessions
Shape Sorter (Xander)	30 min	114 trials	8 sessions
Average	27 min	109 trials	8 sessions
Standard Multielement			
Receptive ID Animals (Bo)	55 min	265 trials	21 sessions
Puzzle (Bo)			
Sort (Hunter)	56 min	236 trials	18 sessions
Puzzle (Hunter)	34 min	134 trials	10 sessions
Block Design (Xander)	49 min	176 trials	13 sessions
Shape Sorter (Xander)	83 min	321 trials	25 sessions
Average	55 min	226 trials	17 sessions
Standard Reversal			
Receptive ID Animals (Bo)	77 min	378 trials	31 sessions
Puzzle – robe, blanket (Bo)	90 min	388 trials	28 sessions
Puzzle – bed, pillow (Bo)	73 min	310 trials	27 sessions
Puzzle – pajamas, book (Bo)	78 min	344 trials	28 sessions
Sort (Hunter)	106 min	404 trials	29 sessions
Puzzle (Hunter)	84 min	344 trials	25 sessions
Block Design (Xander)	88 min	360 trials	26 sessions
Shape Sorter (Xander)	55 min	215 trials	18 sessions
Average	81 min	343 trials	27 sessions

Table 4. Efficiency outcomes for the brief multielement, standard multielement, and standard reversal across each of the tasks in Study 2.

Participant	Correspondence Between Skill Assessment and Extended Evaluation		
	Brief Multielement	Standard Multielement	Standard Reversal
Bo			
Receptive ID Animals	No	No	No
Puzzle	n/a	n/a	Yes (3)
Hunter			
Sort	Yes	Yes	Yes
Puzzle	Yes	Yes	Yes
Xander			
Block Design	Yes	Yes	No
Shape Sorter	No	No	Yes
Total	3/5	4/5	6/8

Table 5. Correspondence between the results of the skill assessment and the results of the extended evaluation for each of the participants in Study 2.
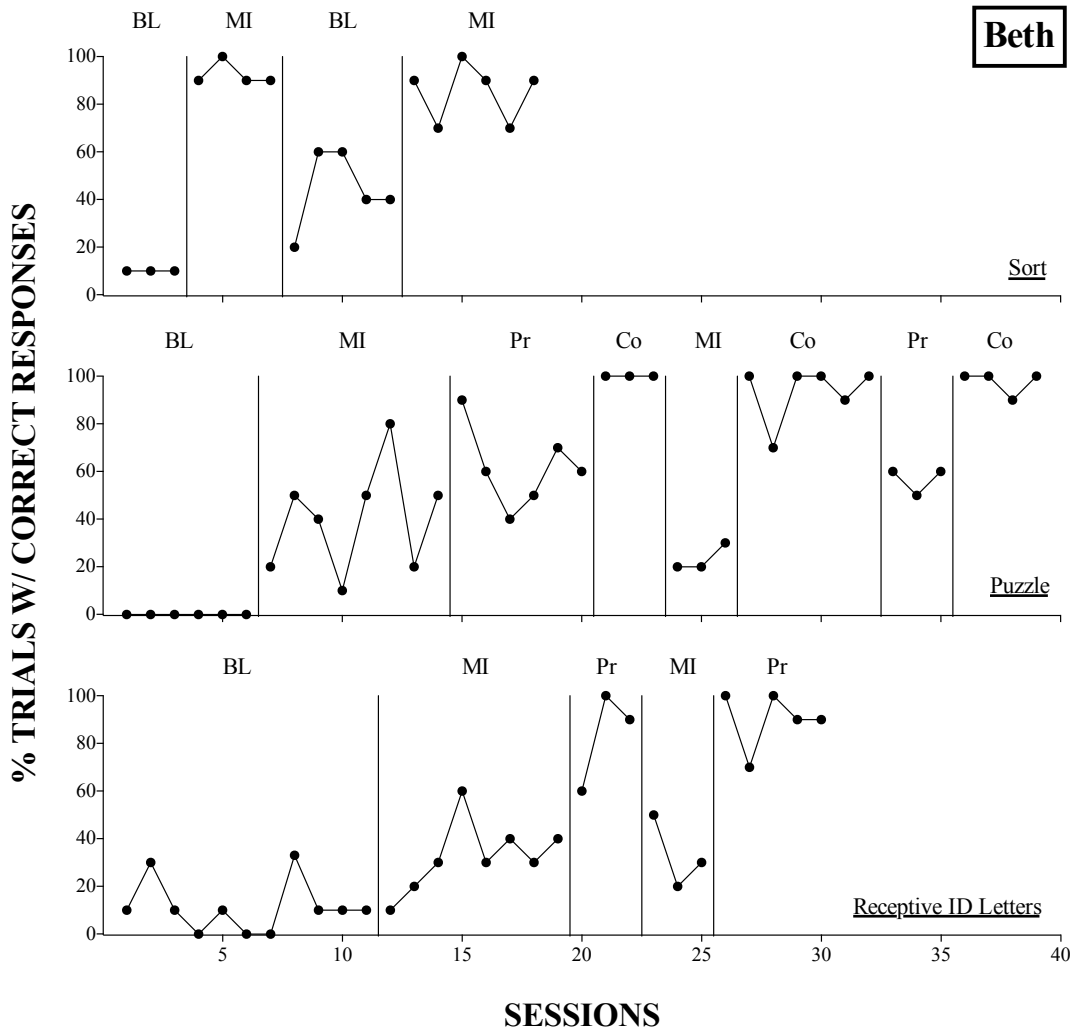
Figure 1. Percentage of trials with correct responses for Beth across all target tasks. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.
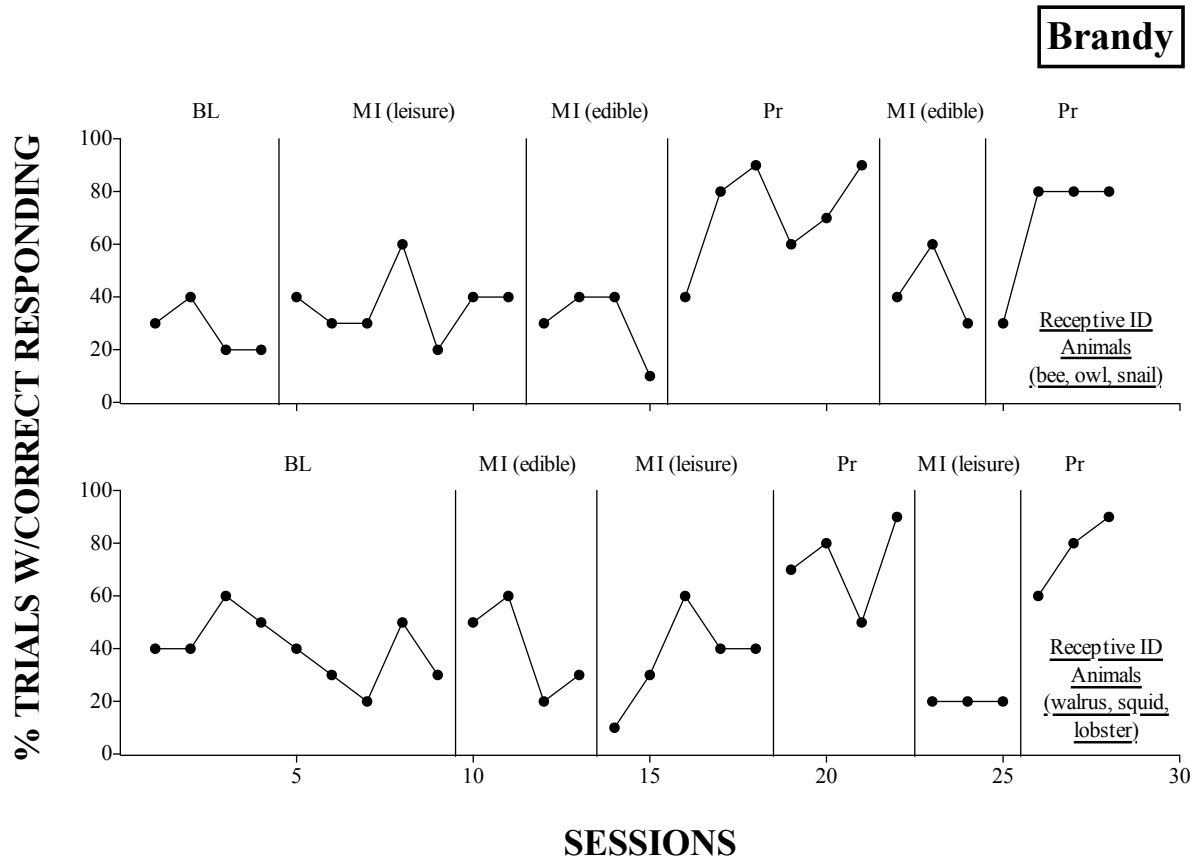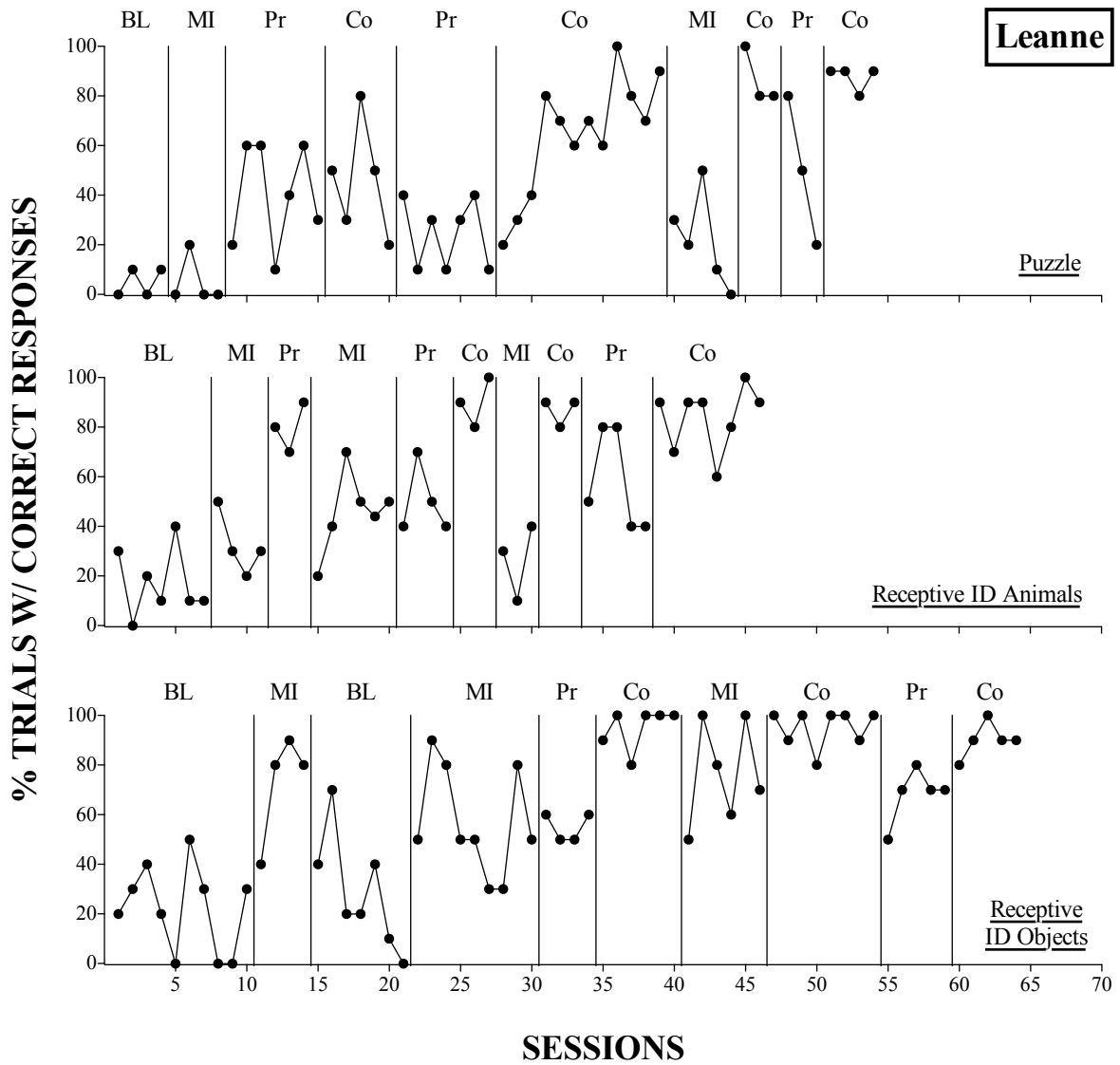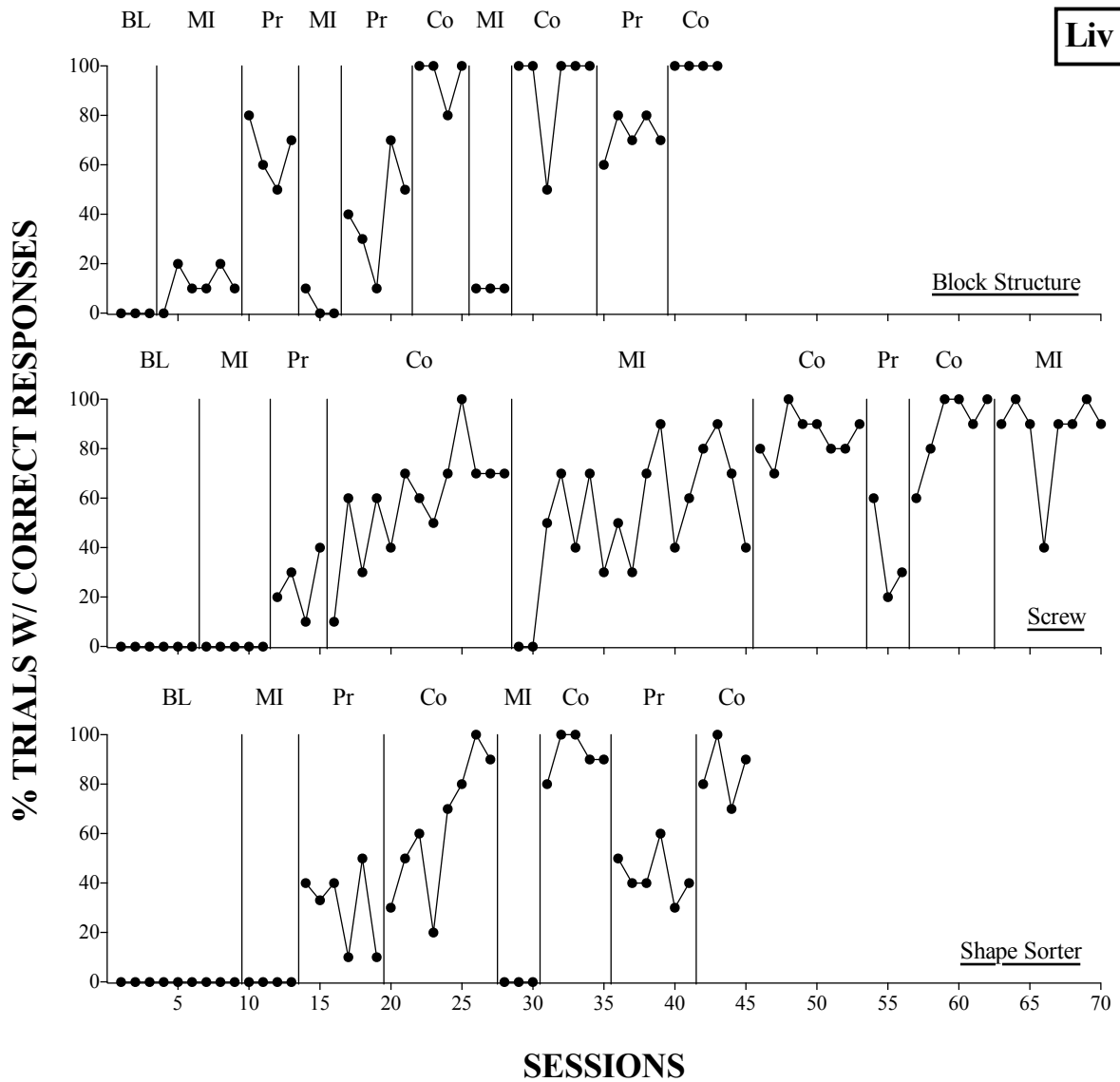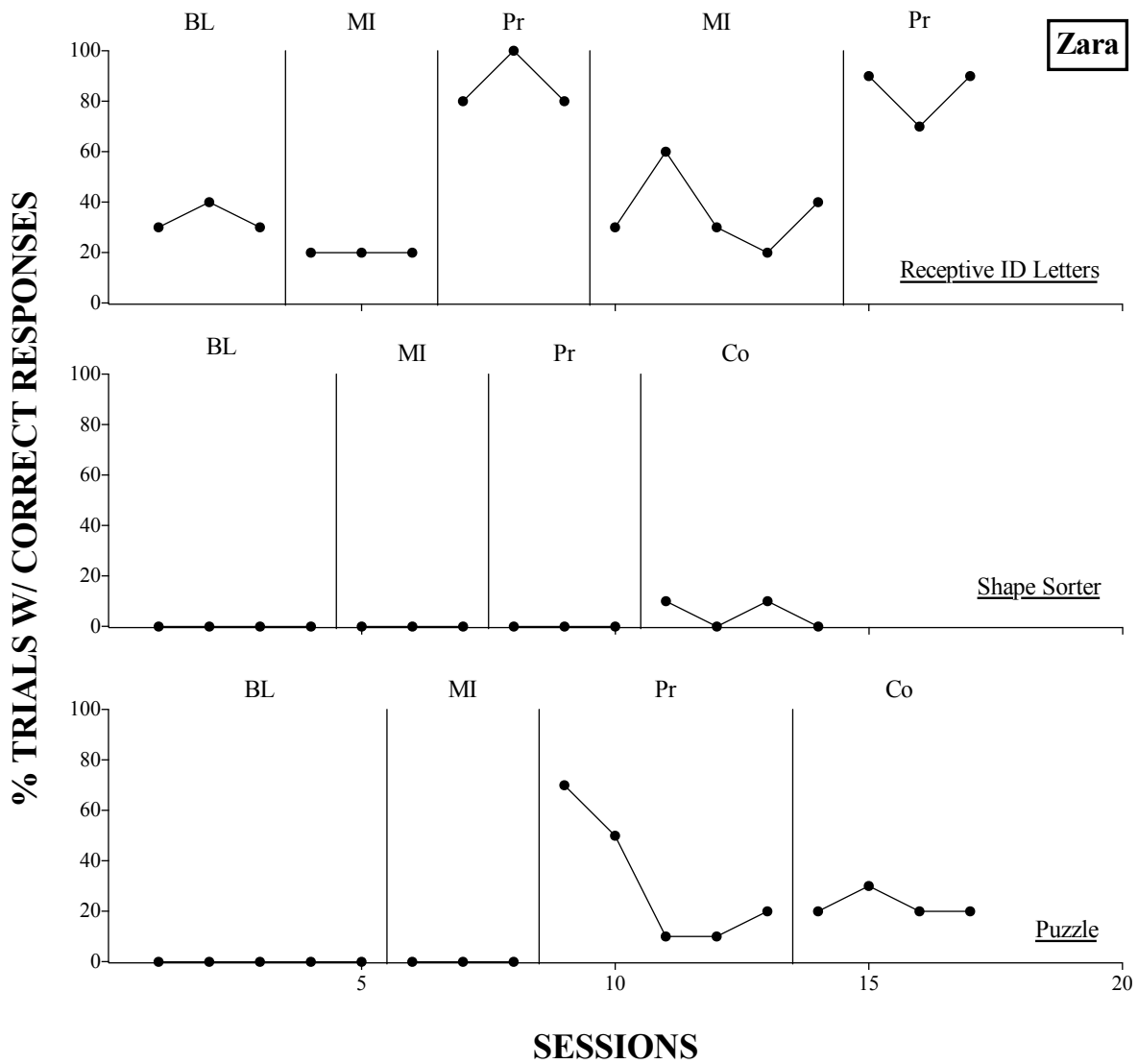
Figure 2. Percentage of trials with correct responses for Brandy across all target tasks. BL = baseline; MI edible = motivational interventions condition using edible items; MI leisure = motivational interventions condition using leisure items; Pr = prompts condition.
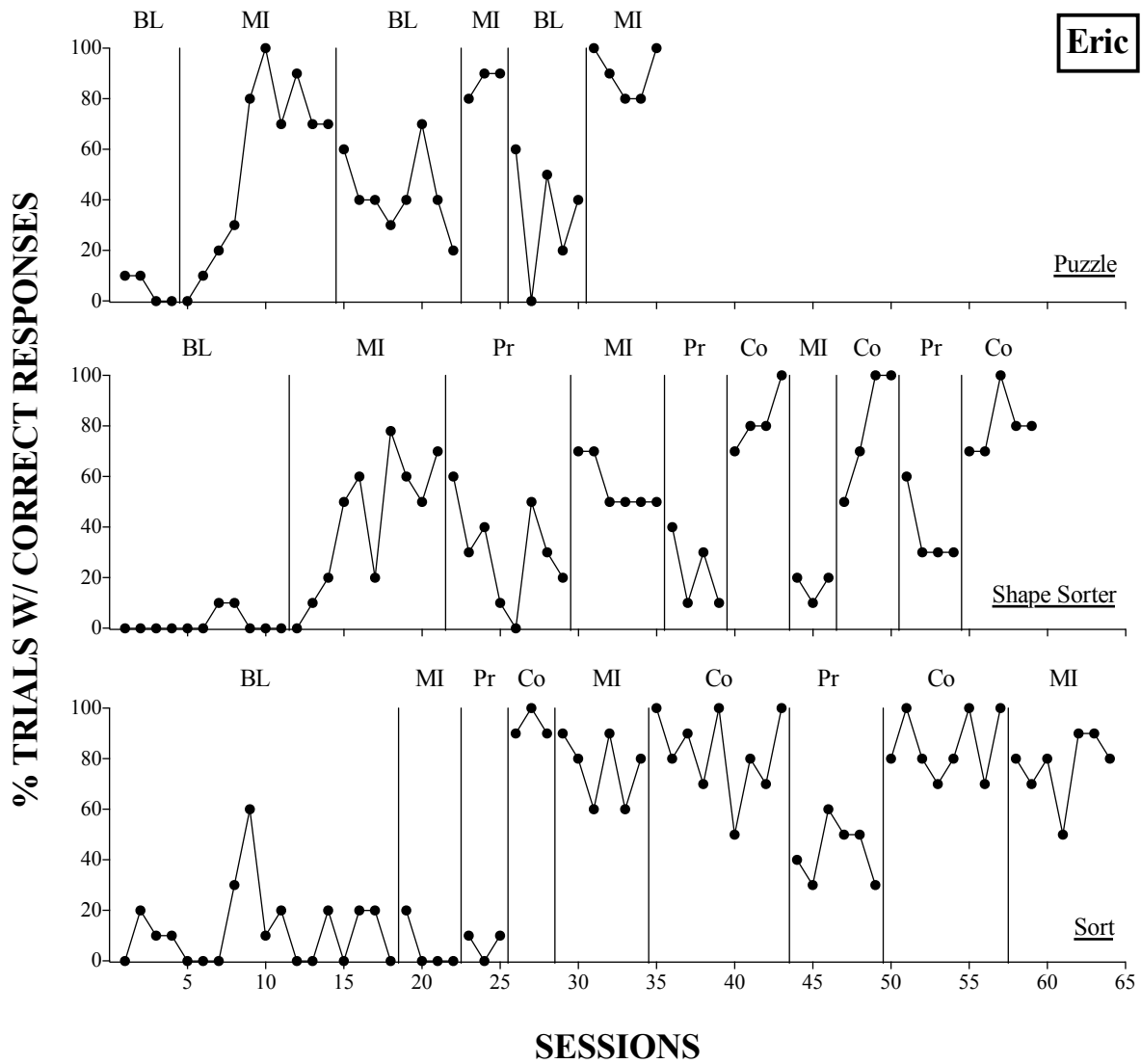
Figure 3. Percentage of trials with correct responses for Eric across all target tasks. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.
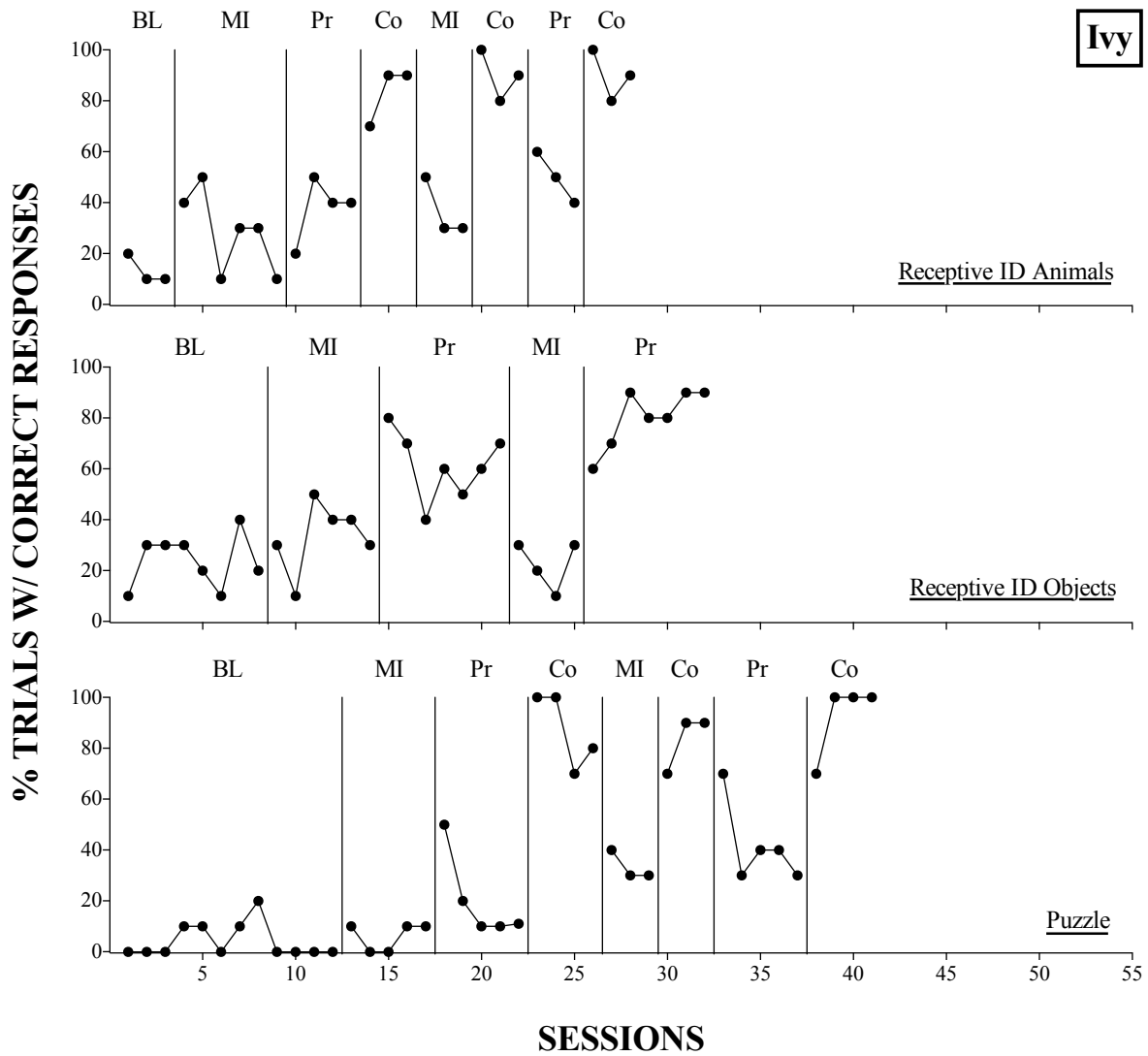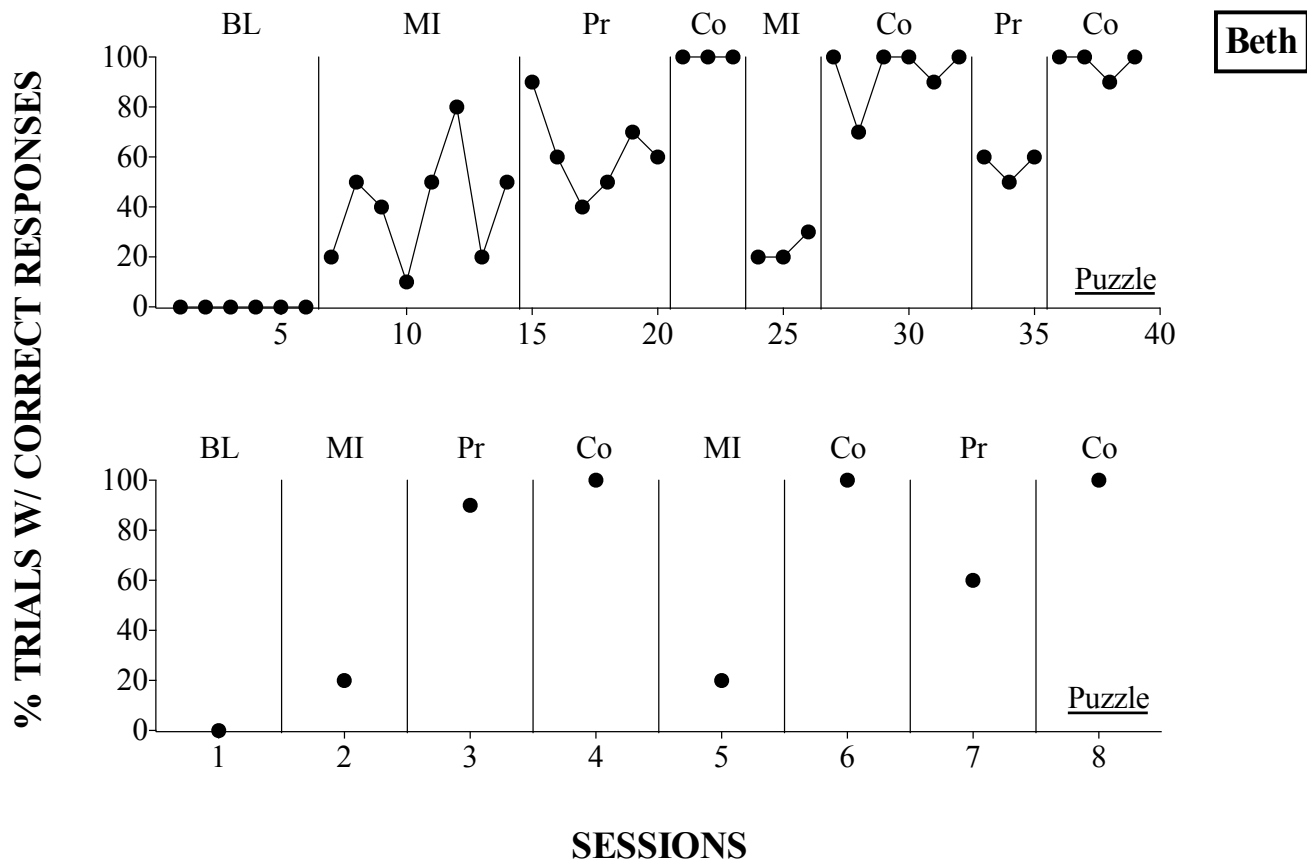
Figure 4. Percentage of trials with correct responses for Ivy across all target tasks. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.
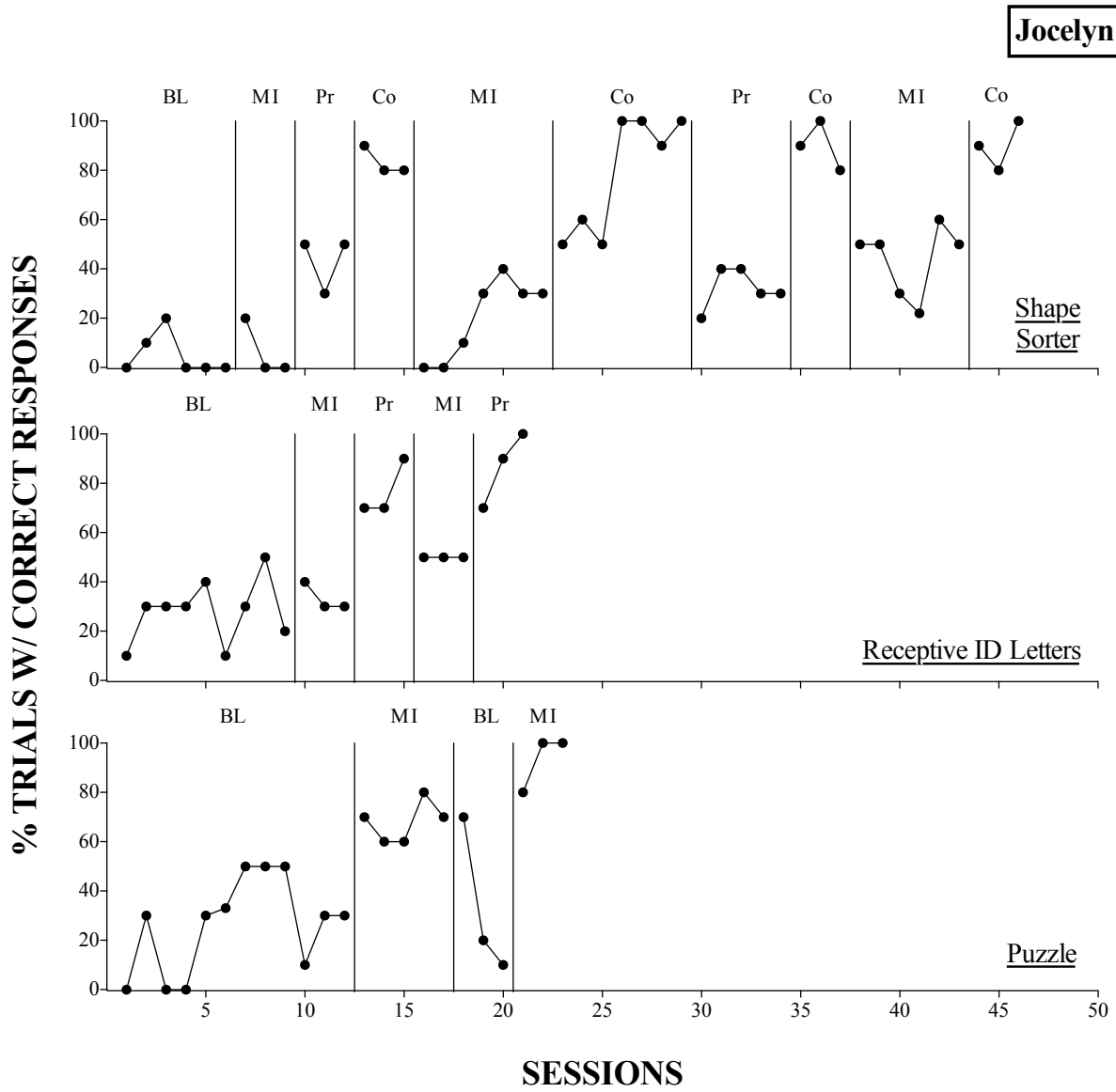
Figure 5. Percentage of trials with correct responses for Jocelyn across all target tasks. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.

Figure 6. Percentage of trials with correct responses for Leanne across all target tasks. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.

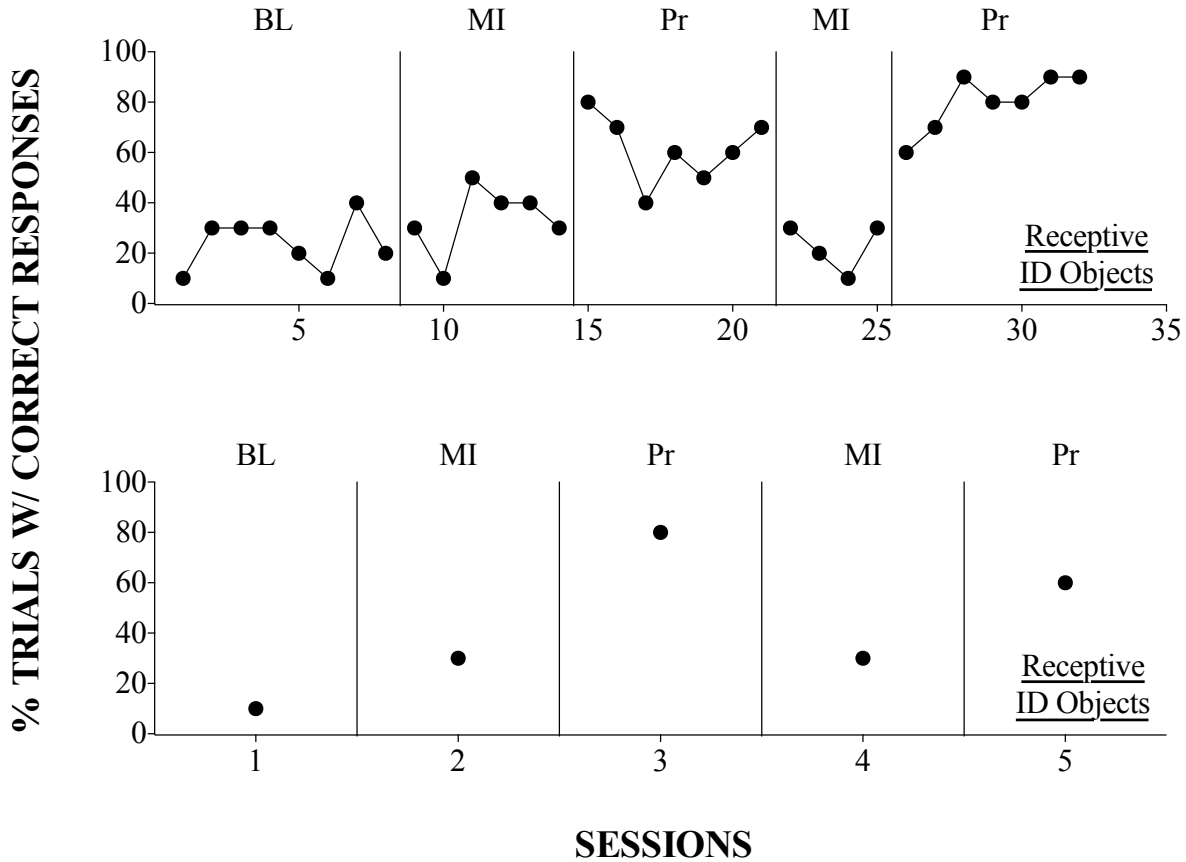*Figure 7.* Percentage of trials with correct responses for Liv across all target tasks. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.

Figure 8. Percentage of trials with correct responses for Zara across all target tasks. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.

Figure 9. Standard depiction of Beth's performance to the puzzle task in Study 1 (top panel) as compared to the single-point analysis (bottom panel) in which Beth's performance to the same task was examined during the initial presentation of the each of the experimental conditions. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.

*Figure 10.* Standard depiction of Ivy's performance to the object identification task in Study 1 (top panel) as compared to the single-point analysis (bottom panel) in which Ivy's performance to the same task was examined during the initial presentation of the each of the experimental conditions. BL = baseline; MI = motivational interventions condition; Pr = prompts condition.

Receptive ID Animals - Bo

Skill Assessment

Extended Evaluation

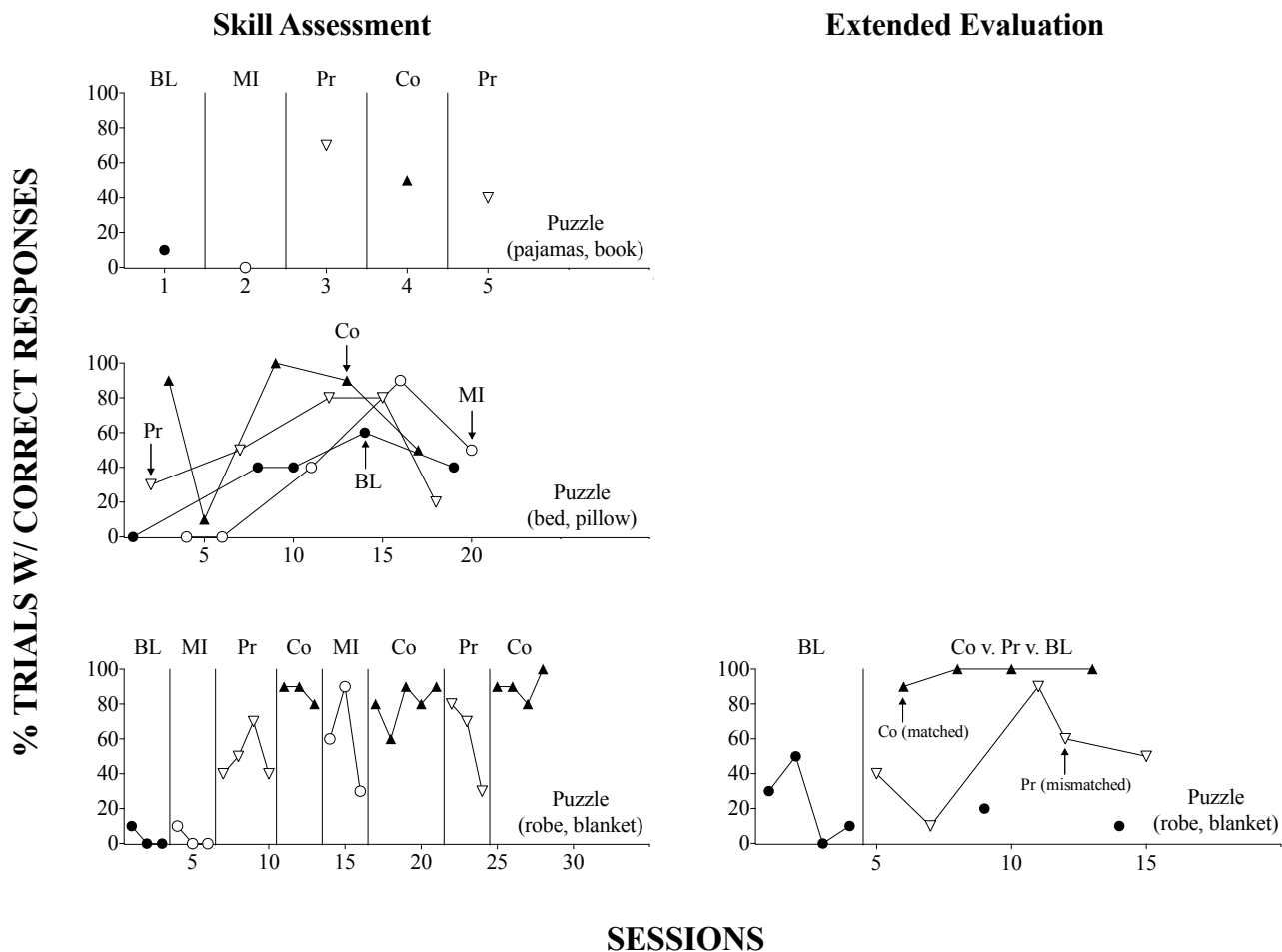% TRIALS W/ CORRECT RESPONSES

SESSIONS
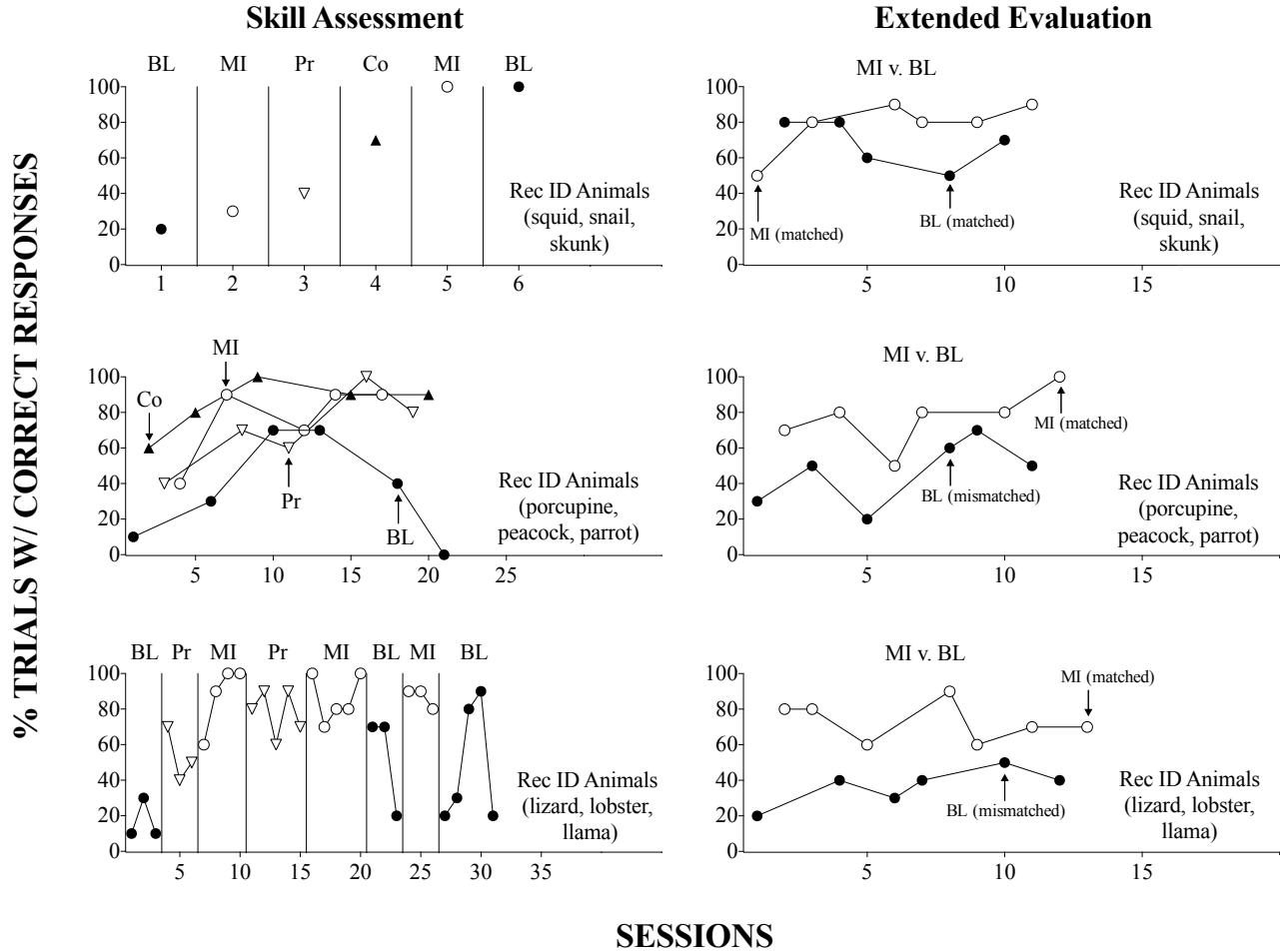
Figure 11. Percentage of trials with correct responses during the brief multielement assessment (top left panel) and corresponding extended evaluation (top right panel), standard multielement assessment (middle left panel) and corresponding extended evaluation (middle right panel), and standard reversal assessment (bottom left panel) and corresponding extended evaluation (bottom right panel) for Bo's receptive identification of animals task. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.

Puzzle - Bo

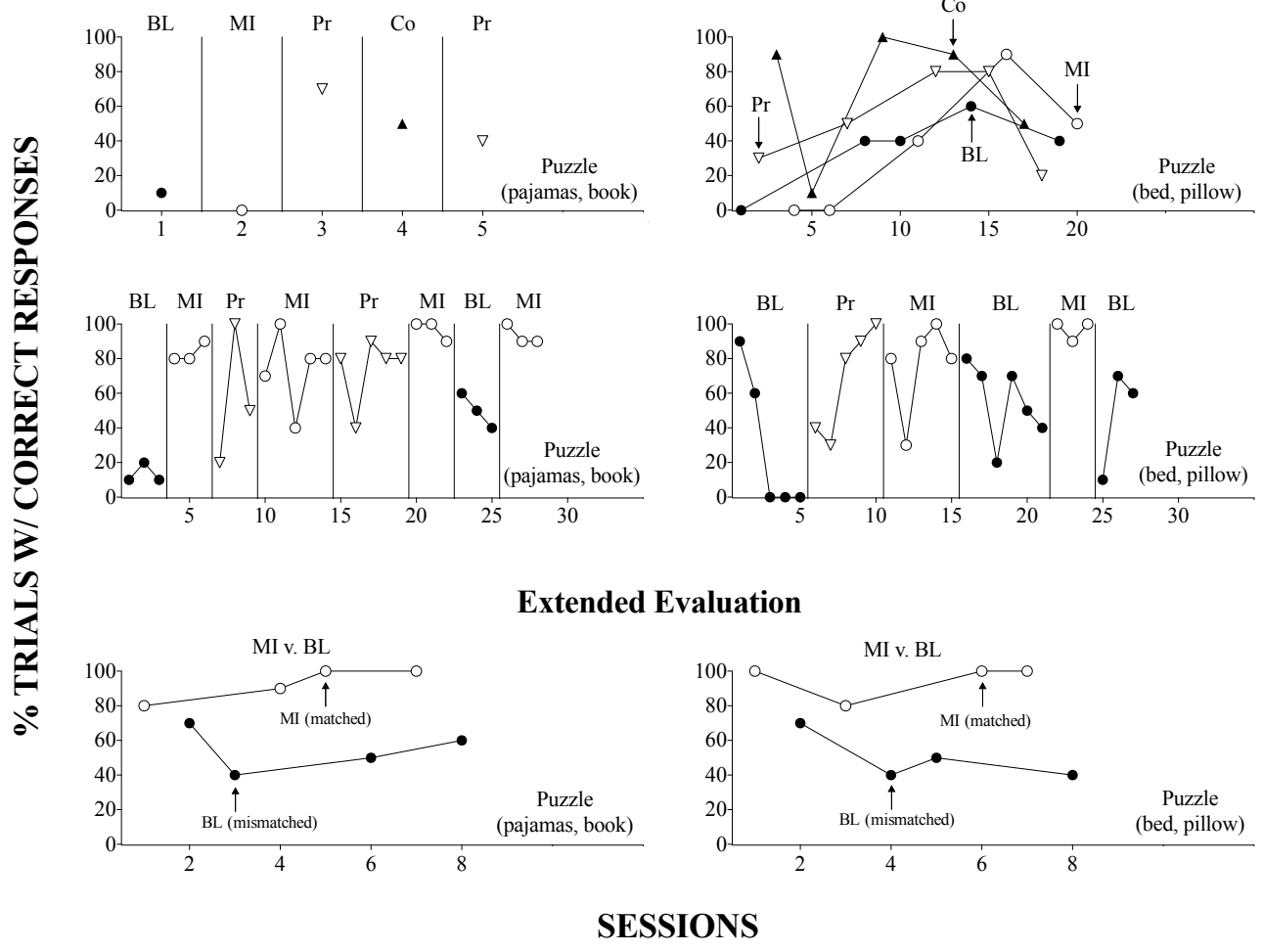*Figure 12.* Percentage of trials with correct responses during the brief multielement assessment (top left panel) and corresponding extended evaluation (top right panel), standard multielement assessment (middle left panel) and corresponding extended evaluation (middle right panel), and standard reversal assessment (bottom left panel) and corresponding extended evaluation (bottom right panel) for Bo's puzzle task. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.

Skill Assessments

Extended Evaluation

SESSIONS

Figure 13. Percentage of trials with correct responses during the brief multielement assessment (top left panel), standard reversal assessment (middle left panel), and corresponding extended evaluation (bottom left panel) for the pajamas and book stimuli for Bo's puzzle task. Percentage of trials with correct responses during the standard multielement assessment (top right panel), standard reversal assessment (middle right panel), and corresponding extended evaluation (bottom right panel) for the bed and pillow stimuli for Bo's puzzle task. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.
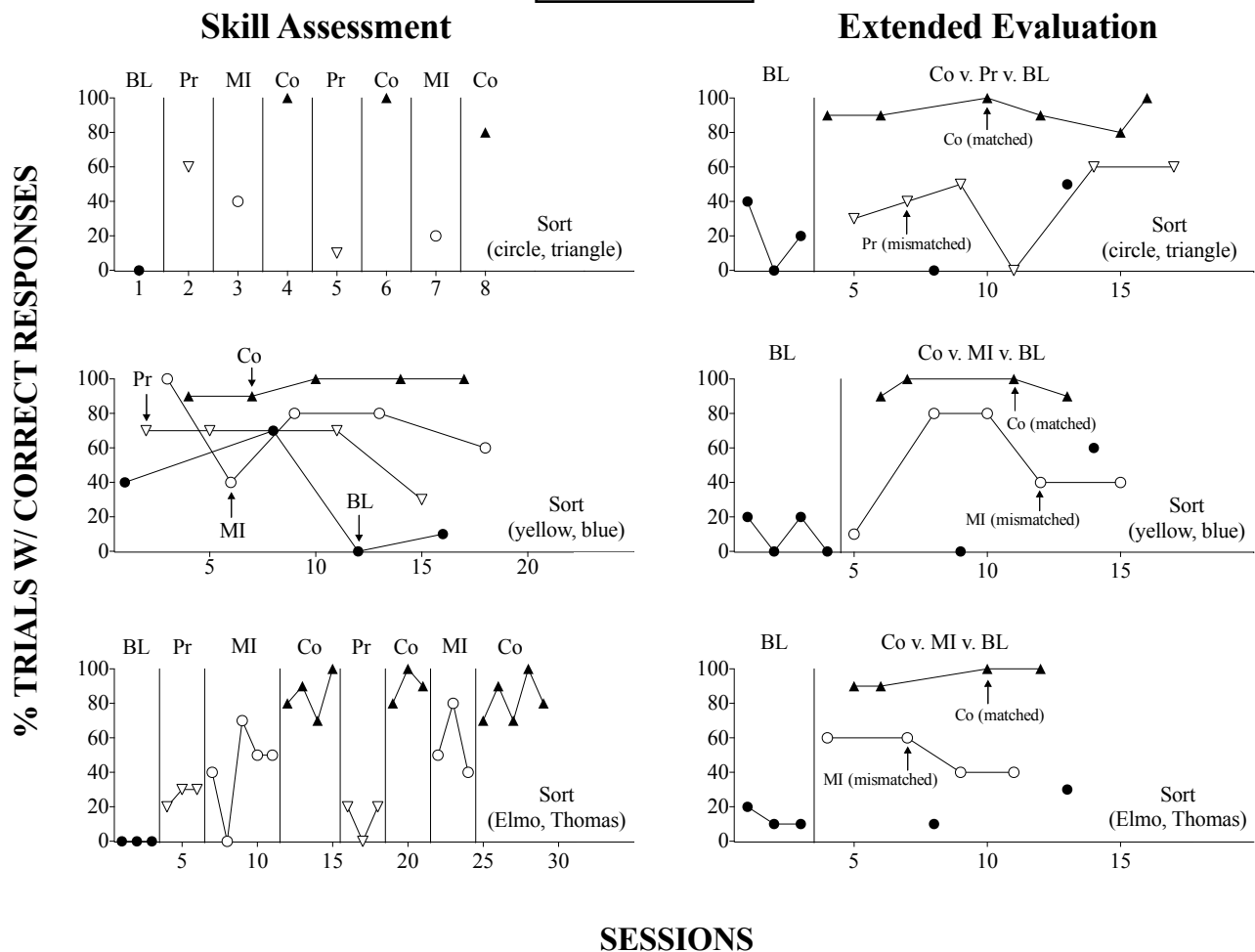
Figure 14. Percentage of trials with correct responses during the brief multielement assessment (top left panel) and corresponding extended evaluation (top right panel), standard multielement assessment (middle left panel) and corresponding extended evaluation (middle right panel), and standard reversal assessment (bottom left panel) and corresponding extended evaluation (bottom right panel) for Hunter's sorting task. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.
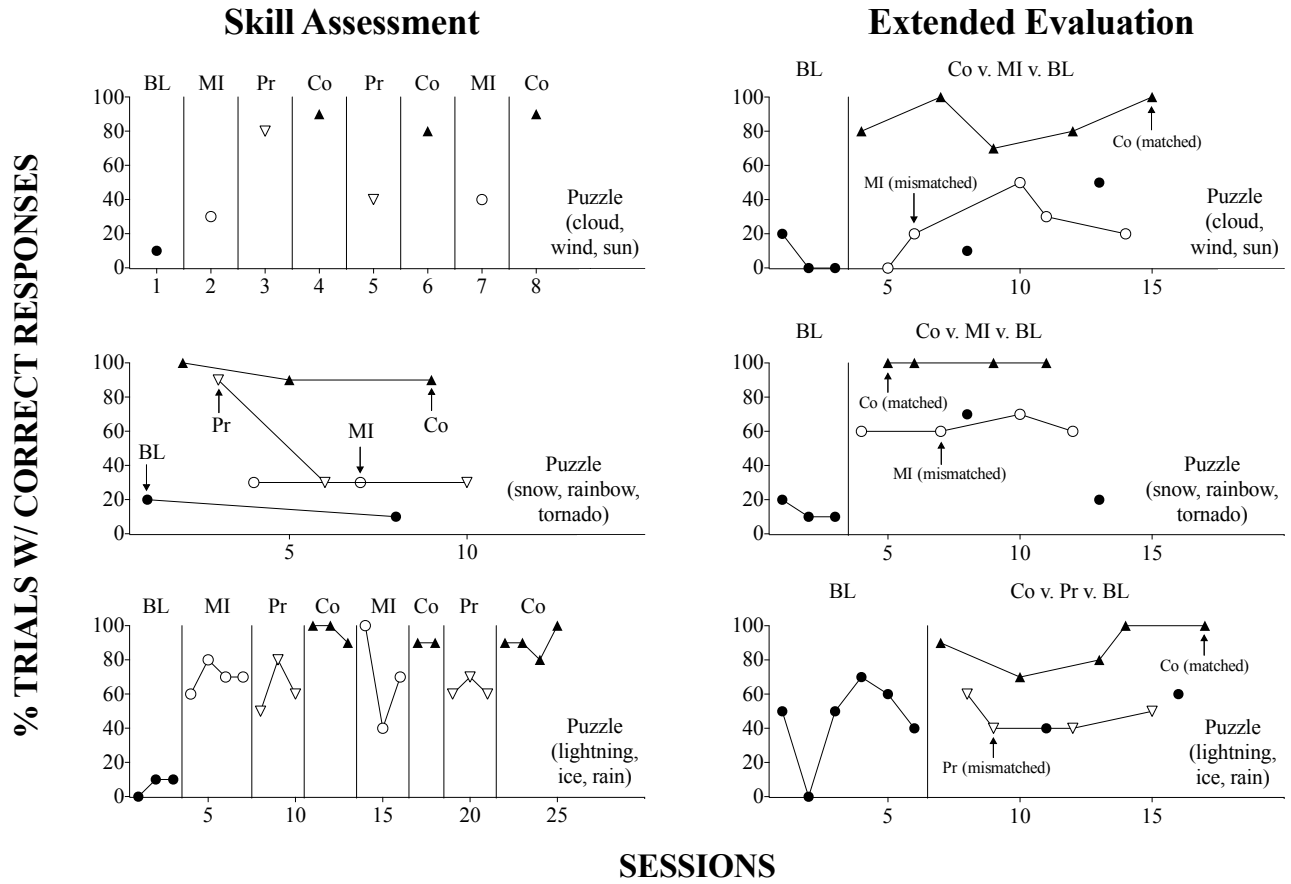
Puzzle - Hunter

Skill Assessment

Extended Evaluation

Figure 15. Percentage of trials with correct responses during the brief multielement assessment (top left panel) and corresponding extended evaluation (top right panel), standard multielement assessment (middle left panel) and corresponding extended evaluation (middle right panel), and standard reversal assessment (bottom left panel) and corresponding extended evaluation (bottom right panel) for Hunter's puzzle task. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.
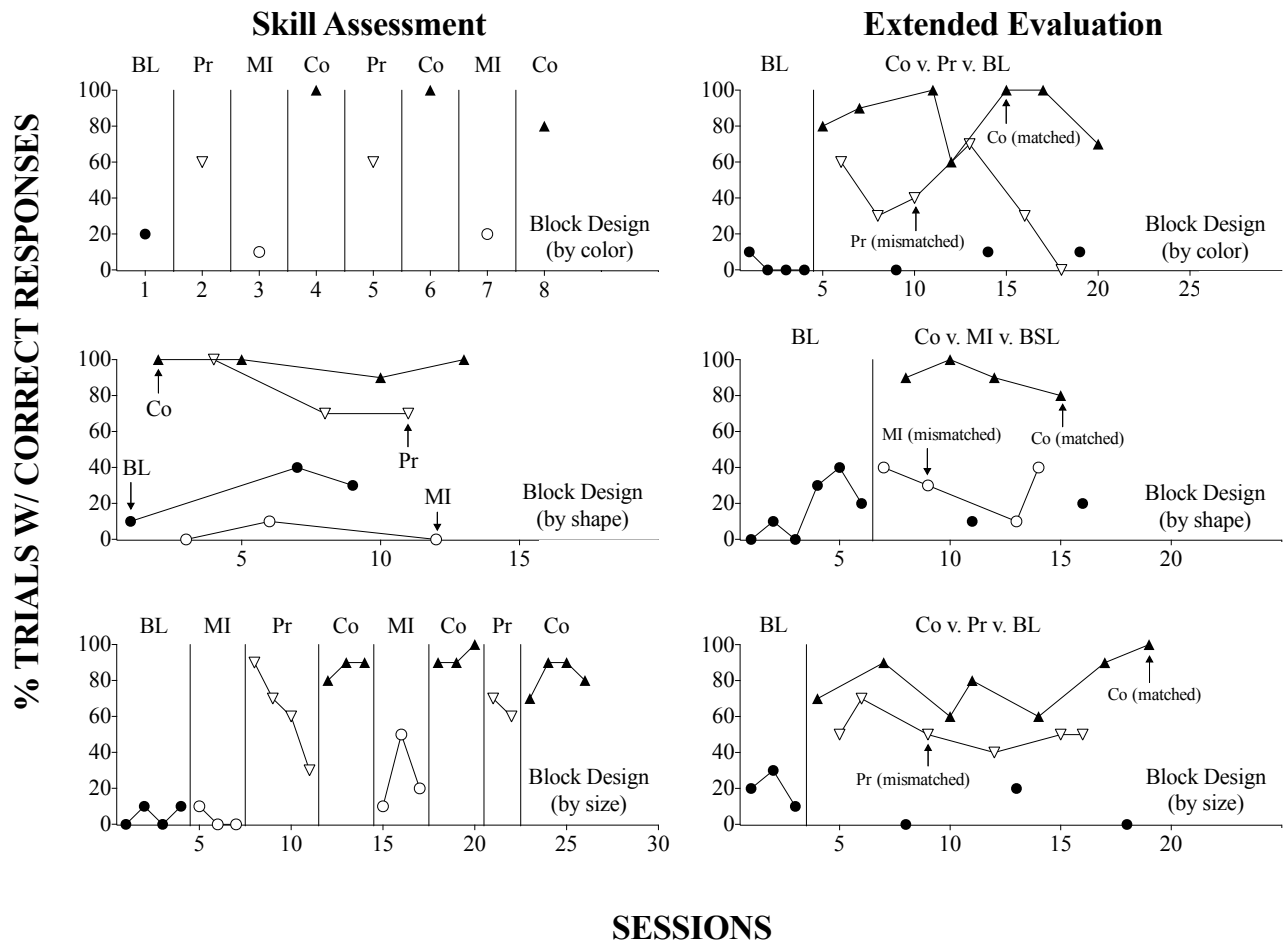
Block Design - Xander

Figure 16. Percentage of trials with correct responses during the brief multielement assessment (top left panel) and corresponding extended evaluation (top right panel), standard multielement assessment (middle left panel) and corresponding extended evaluation (middle right panel), and standard reversal assessment (bottom left panel) and corresponding extended evaluation (bottom right panel) for Xander's block design task. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.
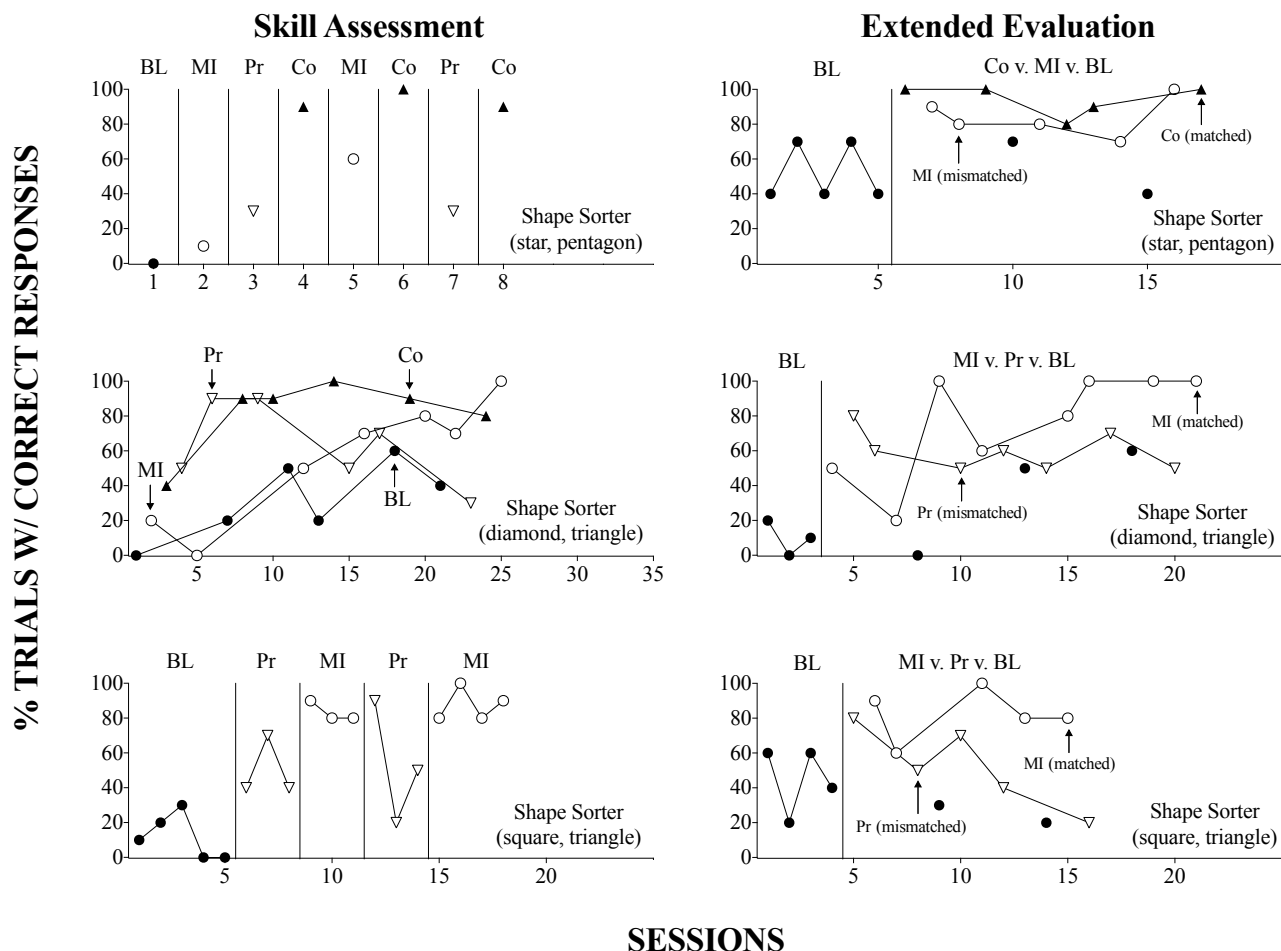
Shape Sorter - Xander

Skill Assessment

Extended Evaluation

% TRIALS W/ CORRECT RESPONSES

SESSIONS

Figure 17. Percentage of trials with correct responses during the brief multielement assessment (top left panel) and corresponding extended evaluation (top right panel), standard multielement assessment (middle left panel) and corresponding extended evaluation (middle right panel), and standard reversal assessment (bottom left panel) and corresponding extended evaluation (bottom right panel) for Xander's shape sorter task. BL = baseline; MI = motivational interventions condition; Pr = prompts condition; Co = combined condition.