



1-1-2017

Role-play assessments as a measure of parenting skill

Scott A. Jensen

University of the Pacific, sjensen@pacific.edu

Carolynn Steiner

University of the Pacific

Follow this and additional works at: <https://scholarlycommons.pacific.edu/cop-facarticles>

 Part of the [Psychology Commons](#)

Recommended Citation

Jensen, S. A., & Steiner, C. (2017). Role-play assessments as a measure of parenting skill. *Behavior Analysis: Research and Practice*, 17(4), 393–401. DOI: 10.1037/bar0000083
<https://scholarlycommons.pacific.edu/cop-facarticles/628>

This Article is brought to you for free and open access by the All Faculty Scholarship at Scholarly Commons. It has been accepted for inclusion in College of the Pacific Faculty Articles by an authorized administrator of Scholarly Commons. For more information, please contact mgibney@pacific.edu.

ROLE-PLAY ASSESSMENTS AS A MEASURE OF PARENTING SKILLS

Abstract

Various forms of assessments of parenting skill are used across studies including parent self-report, observations of parent and child interactions, and role-play assessments of parenting skill. Role-play assessments, though used in several studies, have not been sufficiently evaluated so as to understand their reliability and validity as measures of parenting skill. The current study compared repeated role-play assessments and parent-child interactions across five weeks for four parents. Consistency across role-play assessments was adequate (80%), but the variability of parent behavior with their own child across sessions (59% similarity) led to poor similarity with role-play assessments (63%). Findings suggest promise for the usefulness of role-play assessments, but raise questions about the stability of parent behavior with their own children during in-home observations.

Keywords: Behavioral Parent Training; Assessment; Behavioral Observations; Parenting Skill

Role-play assessment as a measure of parenting skills

Behavioral Parent Training (BPT) is one of the most commonly used interventions for Child problem behaviors, and several programs are considered “Well-Established” and “Probably Efficacious” treatments (Eyberg, Nelson, & Boggs, 2008). These programs target child behavior change by training parents to implement behavioral skills with their children, including appropriate play, praise, tangible rewards, giving effective commands, planned ignoring, time-out, etc. Change in parent behavior is assumed to be the mechanism that leads to child behavior change (Reid & Patterson, 1989).

Most BPT programs are evaluated through two means: parent/teacher report scales and behavior observations. Self-report measures are easy to administer, thus making them the most common evaluation method. Unfortunately, they present several limitations including self-report bias and inaccuracies in reporting (Pelham & Fabiano, 2008), leading both Eyberg et al. (2008) and Pelham & Fabiano (2008) to note the importance of not relying on studies that included only parent/teacher report measures in assessing empirically supported treatments for children and adolescents. As a result, direct observation is considered the gold standard for assessing parent and child behavior change (Patterson, Reid, & Eddy, 2002).

Behavior observation coding systems have been developed to capture multiple behaviors and environmental variables that surround these behaviors (i.e., antecedents and consequences) in a single observation. The most commonly used measure is the Dyadic Parent-Child Interaction Coding System (DPICS; Robinson & Eyberg, 1981), a behavioral coding system that directly observes parent-child interactions and scores the frequency of different domains of parent and child behaviors to examine the quality of interactions. Parent domains include verbalizations (i.e., negative talk, indirect or direct commands, labeled or unlabeled praise, informative

questions, descriptive questions, reflective statements, behavioral descriptions, and neutral talk) and physical responses (i.e., positive or negative touches). Child domains consist of verbalizations (i.e., negative talk, commands, questions, pro-social talk), vocalizations (i.e., yelling, whining), responses (i.e., answer, no answer, no opportunity for answer, compliance, non-compliance, no opportunity for compliance), and physical categories (i.e., positive or negative touches). Parents are instructed on what to do with their child during three situations, (i.e., child-directed play, parent-directed play, and cleanup) with each situation lasting 5 minutes and always occurring in the same order. The coding process is fairly extensive, and unfortunately, the large variety of observed behaviors results in variability in what is reported in published studies, thus creating difficulty in comparing across findings. Other similar observation measures, such as the Intrapersonal Behavior Construct Scale (IBCS; Kogan & Gordon, 1975), the Observations of Parent-Child Interactions (Forehand & McMahon, 1981) and the Family Observation Schedule (Dadds & McHugh, 1992) are also commonly used. These differ in the specifics of how and what is measured, but are additional methods for coding child and parent interactions based on observation.

Parent Behavior Change as a Program Evaluation Measure

While the DPICS and other observation measures allow for global assessment of parent and child behaviors, they do not always assess the specific skills taught in BPT. Consequently, few studies examine changes in specific skills addressed in training. Furthermore, these observations tend to be used as a single pre and post observation. A less commonly used method of measuring improvements in skill acquisition is the use of role-play assessments, in which a trained individual plays the role of the child, thus allowing for control of child behavior. Role-play assessments allow individuals (in this case, parents) to show skill effectiveness in a planned

scenario that provides opportunities to demonstrate all necessary skills. Two distinct advantages of having control over child behavior in role-play assessments of parenting skills are a) decreased assessment time, and b) assessment of response to less frequent behaviors (especially disruptive behavior).

Direct observations, in the form of role-play assessments to measure specific skills, have been shown to be useful in several studies of parenting skill, though no direct evaluations of the measurement tool itself have been published. Cobb, Leitenberg, and Burchard, (1982) attempting to evaluate a parenting program based on skill acquisition of the parents, demonstrated that role-plays can be useful in measuring parent behavior. Van Camp et al. (2008) also utilized role-play assessments in parent observations with an adult role-player to assess parent skill acquisition in a two-part study demonstrating the success of their program. Berard and Smith (2008) similarly demonstrated the use of role-plays when evaluating a modified version of the same parenting program. Across participants, role-play scores averaged 38.55% during baseline and increased to an average of 71.87% post-treatment. Two limitations stood out from the use of role-plays in these studies: 1) observers were not blind to training status, and 2) no comparisons were made between behavior with the role-play “child” and the actual child to insure similarity.

The purpose of the current research was to examine the reliability and validity of role-play assessments of parenting skill by comparing repeated measurements and addressing previous limitations. Improving on previous studies, observers were blind to the training status of the participants, decreasing the likelihood of observer expectancies. Also, parents were not informed what specific skill was being assessed through the role-play (Berard et al., 2008; Van Camp et al., 2008). Similar performances across both types of observations would help validate

the use of role-play assessments by suggesting minimal differences between using an adult role-player compared to the parent's child. Given the advantages of control and time efficiency with an adult role-player, this finding would be important.

Method

Participants were four parents with children ranging in age from 2 to 4 years.

Participants were recruited through preschools and word of mouth from other participants. They were included based on parent expression that they were concerned about their child's behavior, but not formal assessment or diagnostic procedures were conducted. To be included, participants had to a) score a 60% or below across two or more skill areas during the first two role-play observations, and b) not be participating in any parenting courses. Each participant was paid \$50 (i.e., \$10 per session) for their participation, and each session was conducted in the participant's home.

A multiple probe design with a single-skill treatment phase was utilized to assess parent behavior with an adult role-player compared to their own child, with the order of conditions counter-balanced across participants. The parent-child and role-play assessments were conducted a total of five times (once per week), with each assessment lasting approximately 30 min. Just before the fourth observation, training was conducted for the skill area in which the participant performed lowest across the first three observations.

Role-play Assessments

Role-play assessments were developed to assess specific skills taught within the Incredible Years (IY) curriculum (Webster-Stratton, 2001), though it should be noted that they are common to most BPT programs. The IY curriculum was being implemented by the researchers and thus the role-play assessments were developed as a method to assess the

outcomes of the program. It is believed that these task analyses could be readily adapted to other specific programs. For the current study, four skills were assessed: Play, Praise, Effective Commands, and Ignoring. Though role-play assessments have been created to assess time-out and use of tangible rewards, they are not included in the present study because of the difficulty in easily assessing such behaviors through comparison parent-child interaction assessment: tangible rewards (required too much time) or time-out (cannot insure child behaves inappropriately). These realities further highlight the usefulness of role-play assessments of parenting behavior. Time-out comparisons have been carried out in other research (Jensen, Blumberg, & Doerr, 2017).

A trained graduate student played the role of the child – referred to below as the role-play “child” – across all skills and participants. In each case, they followed a specific script of behaviors that provided opportunities for parents to demonstrate each of the skills within the task. The specific scenario was explained to the parent, who was then instructed to respond as they would with their own child. Which skill being assessed was not made clear, and no feedback was provided about the role-plays or the parent-child interactions. All assessments were video recorded for later data collection. Integrity checks on role-player behavior were conducted for all role-plays to ensure consistency across sessions and participants.

Semi-structured Parent-child Interactions

Semi-structured parent-child interactions were organized into four sections to match the four role-play assessments. Each participant was a) told to play with their child for 5 minutes (Play), b) told to do a coloring or writing activity with their child (5 min- Praise), c) asked to tell their child to play by themselves while the parents worked (5 min - Ignoring), and d) asked to give their child 3 commands, as part of having them clean up the toys (Commands).

Improvement was defined as a two-or-more-step improvement within a skill. Scores that were within one step above or below were considered to have not changed. See Table 1 for the specific skills in each task analysis. The task analyses were created by listing the components of each skill identified as a key to the overall skill within the curriculum. This was accomplished by the authors through mutual agreement, based on a review of the parent book and instructor materials.

Two undergraduate data collectors were trained to score the role-plays and parent-child interactions via task analysis for each skill area. Each step within a skill was marked as correct, incorrect, or N/A if the opportunity did not occur during the observation. Total number of steps correct was then divided by steps possible to create a percentage of parent accuracy. Data collectors achieved 85% inter-observer agreement (IOA) with a master coder prior to coding for the study, and each had at least one-year experience coding such videos. Videos were blind coded, - without awareness of which session was being coded. Agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements, and multiplying by 100 to yield a percentage. IOA was collected for 96% of all role-plays and semi-structured observations. Total IOA across all assessments was 86%.

As is customary in single-subject/repeated measures designs, visual analysis was used to compare role-play and parent-child interactions, as well as to examine change in performance across skill areas following training.

Skill Training

To determine which skill to train for each participant, the main experimenter coded the role-plays for the first three observations, and determined the skill area that showed the lowest percent of steps correct across the first three role-play observations. For the actual analysis,

however, two trained observers, blind to session order and skill used for training, coded all sessions. Brief formal training on the skill chosen was provided just prior to the fourth observation. This training lasted approximately 30 min, with the teacher following a checklist of what to cover for the selected skill area. The teacher showed a brief video vignette from the Incredible Years curriculum (Webster-Stratton, 2001), didactically discussed the skill steps with the participant, and concluded by facilitating a brief practice. Immediately following the training, the parent participated in the role-play and parent-child assessments just as before.

Results

Graphs of role-play observations and parent-child interactions across four skills can be found in Figure 1. While some behaviors improved between the first and second baseline, these improvements did not persist in most cases, and seemed to reflect more the variability of behavior over time rather than practice effects or reactivity. Overall, accuracy was moderate, suggesting that floor and ceiling effects would not negatively impact measures of similarity and reliability (see Table 2 for percentages of accuracy, similarity, and reliability within and across scores).

Similarity of Role-Play and Parent-child Interactions

Overall, scores between the role-play assessments and the parent-child assessments were similar, though not strongly consistent. When looking at similarity (within one step), similarity was moderate for three of the four participants (P1 = 71%, P3 = 72%, P4 = 71%, and P2 = 39%; average = 63%). This moderate similarity was strongly affected by the overall variability of parent behavior across the observations. In fact, when using the same criteria of within-one-step to compare parent accuracy across repeated parent-child interactions only, average similarity was even lower (P1 = 69%, P2 = 57%, P3 = 79%, and P4 = 38%; average 59%). Reliability within

role-play assessments alone was 80% across all four participants, being on the lower end of acceptable reliability. Similarity between role-play assessments and parent-child interactions varied by skill, with Play being the lowest (40% similarity), and Praise being the highest (75% similarity). Similarity for Effective Commands was 61% and for Ignoring was 69%. Reliability also varied within parent-child interactions as well, though not showing the same trend as the variability between assessments (Play = 63%, Praise = 56%, Effective Commands = 79% Ignoring = 33%).

Response to Training

The lowest skill area for each participant based on the first three observations was chosen for training. Ignoring was selected for P1, P3, & P4 and Commands was selected for P2. Improvements in the role-play observations immediately following training were observed for P2, P3, & P4. These increases ranged between two and four steps, suggesting training had a positive effect on role-play observations. For parent-child interactions, however, smaller changes were noted (0 to 2 steps) immediately following training. A one step increase was seen for P2 & P4 between sessions four and five, but a one-step decrease for P3 during the same time. For P1, no opportunities for Ignoring occurred during the parent-child interactions following training, and so no effect could be determined.

Discussion

Reliability for the role-play assessments was within the acceptable range (80%). Reliability within the parent-child interactions, however, was much lower, at 59% similarity. This greater variability within the parent-child interaction assessments led to lower overall similarity between the role-play assessments and parent-child interaction assessments (63%). Variability across all skill areas were observed with no indication that parents performed

consistently better or worse with an adult role-player in comparison to with their own child. Similarity was somewhat better for the skill of Praise and somewhat worse for Play. Thus, while role-play assessments demonstrated acceptable reliability and better consistency with parent-child interactions than those interactions with themselves, strong conclusions on its validity cannot be drawn from the present data. Related research on role-play assessment for time out, has demonstrated stronger evidence for similarity between role-play and parent-child interactions (Jensen, Blumberg, & Doerr, 2017).

While the primary purpose of the present study was to evaluate the usefulness of a set of role-play assessments for measuring specific parenting skills, perhaps the most notable finding is the overall variability of parent behavior across time with their own child. The overall reliability of observations from parent-child interactions across four skills during five weekly observations was 59%. This suggests parent behavior varies significantly across multiple observations. It calls into question the common practice of taking a single pre and post observation as a measure of treatment improvement, as this may not be a reliable assessment of parent behavior. All of the BPT programs determined to be “Well-Established” or “Probably Efficacious” by Eyberg, Nelson, and Boggs (2008) used only a single pre and single post observation as this is almost a universal practice amongst those that even use observation (most use only parent report measures). The current findings suggest that there is likely a significant level of error in single observations given the variability of parenting skill implementation over multiple observations. Even in studies of Parent-Child Interaction Therapy, where multiple assessments are made across all sessions as part of treatment, published studies only report a single pre and post observation (Nixon, Sweeney, Erickson, & Touyz, 2003; Schuhmann, Foote, Eyber, Boggs, & Algina, 1998). While these findings of relatively poor stability of parent skill implementation over time are only

preliminary and further evidence should be gathered, they do suggest caution in interpreting a single observation of parent behavior.

One confound that may contribute to variability in parent-child interactions is that not all skill steps or scenarios were applicable during the parent-child interactions, causing the total steps possible to vary across observations. Another variable that may have led to a lack of similarity between the role-plays and the parent-child interactions is the saliency of the cues that might evoke parent behavior. The role-player engaged in behaviors that may have been more severe than what the parent is used to with their own child. These differences may have contributed further to variability throughout the observations.

In a way, these findings make the point further for why role-play assessments have potential value. They allow for control over the child behavior, such that appropriate parent behavior can be evoked in a controlled, shorter period of time. They may, in fact, allow for greater stability of demonstrated parenting skills. Role-plays do possess a certain type of internal validity, given that they are developed based on exactly what is taught and that opportunities to engage in skill steps are evoked by the role-player.

In terms of sensitivity to training, the role-play assessments show promise. Role-play assessments detected improvement immediately following treatment for three of the four participants, which was similar to the results Cobb et al. (1982) and Van Camp et al. (2008) experienced. Scores for parent-child interactions, however, did not demonstrate much improvement. Of the three participants with scores for the parent-child interactions, improvements in the parent-child interactions were observed only during the final session for two participants. The third participant showed improvement immediately following treatment; however, there was already an increase in performance during the final baseline session. The

data from the parent-child interactions make it unclear if the increases observed were due to treatment or a practice effect. It is also unclear how the trends may have continued if data points had been collected over a longer period of time following training. It may be that improvement in parent-child interactions may have matched gains on role-plays, or that initial gains in role-plays may have lessened over time. Longer post-training data collection would help to clarify this.

In conclusion, the use of role-plays that are tailored to the specific skill steps taught in a parenting course are potentially useful in determining skill acquisition given that all skill steps can be evoked by an adult role-player. However, the results are mixed on how well parent behavior with an adult role-player generalizes to scenarios with the parent's own child. Since this study only included three baseline data points, future research could look at extending the baselines until stabilization occurs to see if there is clear improvement immediately following training. By extending data collection following treatment, future research could clarify if improvements in the structured scenarios maintain over time and generalize to the child scenarios. Further research is also needed to replicate the variability of parent behavior with their children across sessions as well as to validate the usefulness of role-play assessments as a measure of parenting skill.

References

- Berard, K. P., & Smith, R. G. (2008). Evaluating a positive parenting curriculum package: An analysis of the acquisition of key skills. *Research on Social Work Practice, 18*, 442-452. doi:10.1177/1049731508318656
- Cobb, E. J., Leitenberg, H., Burchard, J. D. (1982). Foster parents teaching foster parents: Communication and conflict-resolution skills training. *Journal of Community Psychology, 10*, 240-249. doi: 10.1002/1520-6629(198207)10:3
- Dadds, M. R., & McHugh, T. A. (1992). Social support and treatment outcome in behavioral family therapy for child conduct problems. *Journal of Consulting and Clinical Psychology, 60*, 252–259.
- Eyberg, S. M., Nelson, M. M., & Boggs, S. R. (2008). Evidence-based psychosocial treatments for children and adolescents with disruptive behavior. *Journal of Clinical Child & Adolescent Psychology, 37*, 215-237. doi: 10.1080/15374410701820117
- Forehand, R. L., & McMahon, R. J. (1981). *Helping the noncompliant child*. New York: Guilford.
- Jensen, S.A., Blumberg, S., & Doerr, M. (2017). Structured feedback for time-out training: Efficacy and efficiency in comparison to a didactic method. Manuscript submitted for publication.
- Kogan, K. L., & Gordon, B. N. (1975). Interpersonal behavior constructs: A revised approach to defining dyadic interaction styles. *Psychological Reports, 36*, 835-846.
- Nixon, R. D. V., Sweeney, L., Erickson, D. B.,^[L]_[SEP] & Touyz, S. W. (2003). Parent-child interaction therapy: A comparison of standard and abbreviated treatments for oppositional defiant preschoolers. *Journal of Consulting and Clinical Psychology, 71*, 251–260. doi:10.1037/0022-006X.71.2.251.

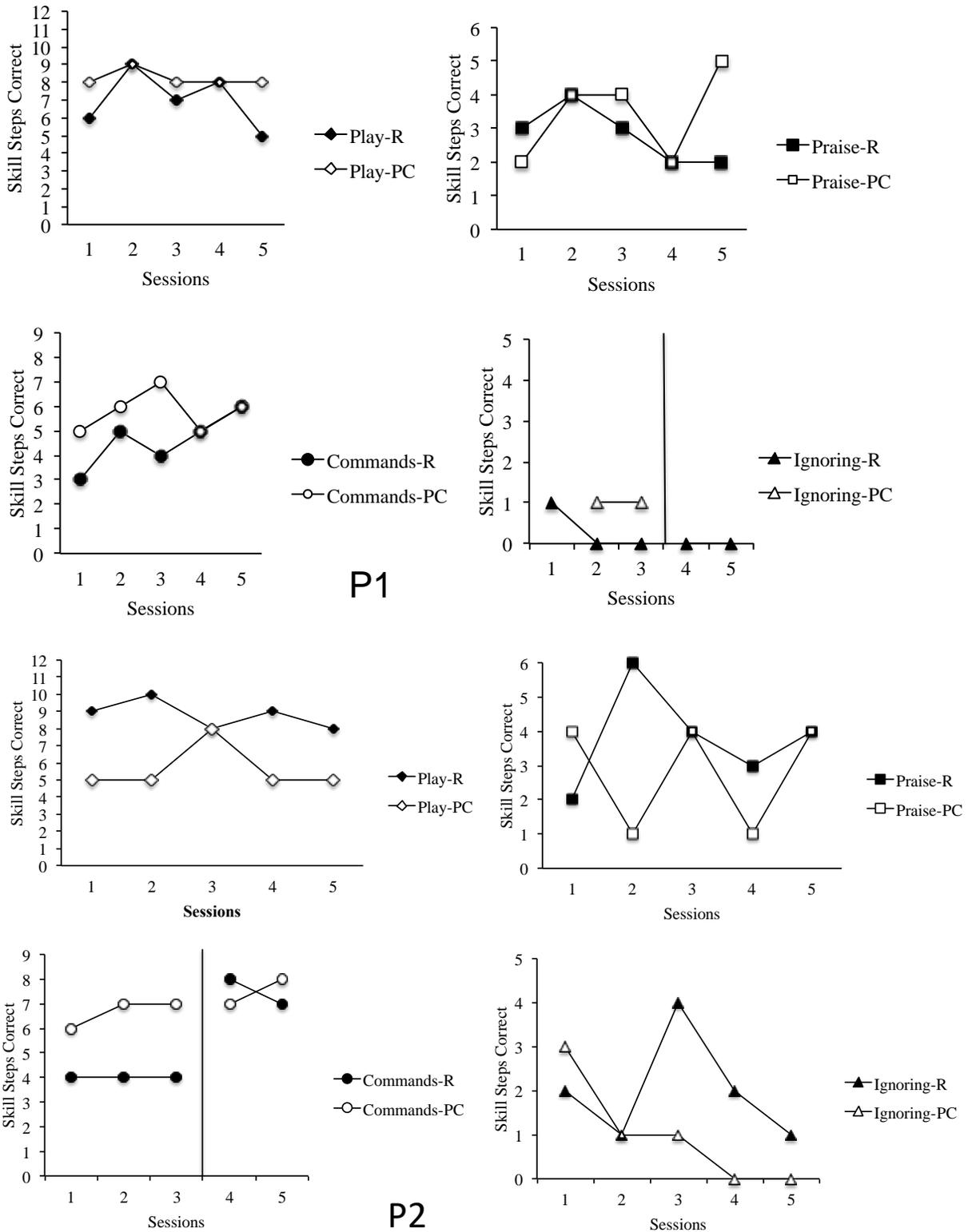
- Patterson, G. R., Reid, J. B., & Eddy, J. M. (2002). A brief history of the Oregon model. In J. B. Reid, G. R. Patterson, J. Snyder (Eds.) , *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention* (pp. 3-20). Washington, DC, US: American Psychological Association. doi:10.1037/10468-001
- Pelham, W. E., & Fabiano, G. A., (2008). Evidence-based psychosocial treatments for attention/deficit hyperactivity disorder. *Journal of Clinical Child & Adolescent Psychology*, 37, 184-214. doi:10.1080/15374410701818681
- Reid, J.B. & Patterson, G.R. (1989). The development of antisocial behavior patterns in childhood and adolescence. *European Journal of Personality*, 3, 107-119.
- Robinson, E. A., & Eyberg, S. M. (1981). The dyadic parent-child interaction coding system: Standardization and validation. *Journal of Consulting and Clinical Psychology*, 49, 245-250. doi: 10.1037/0022-006X.49.2.245
- Schuhmann, E. M., Foote, R. C., Eyberg, S. M.,^[1]_[SEP] & Boggs, S. R. (1998). Efficacy of parent-child interaction therapy: Interim report of a randomized trial with short-term maintenance. *Journal of Clinical Child Psychology*, 27, 34-45.
doi:10.1207/s15374424jccp2701_4.
- Van Camp, C. M., Vollmer, T. R., Goh, H., Whitehouse, C. M., Reyes, J., Montgomery, J. L., & Borrero, J. C. (2008). Behavioral Parent training in child welfare: Evaluations of skill acquisition. *Research on Social Work Practice*, 18, 377-391.
doi: 10.1177/1049731507314008
- Webster-Stratton, C. (2001). *The Incredible Years Teacher Training series*. Seattle, WA: Incredible Years.

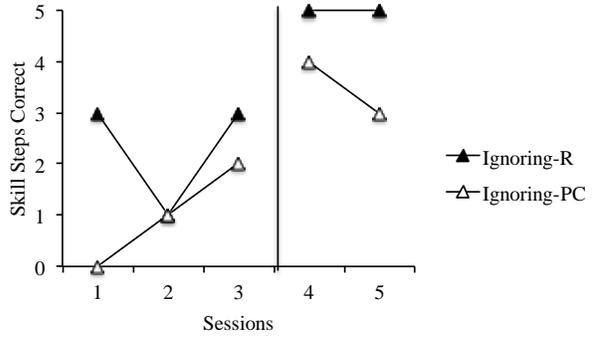
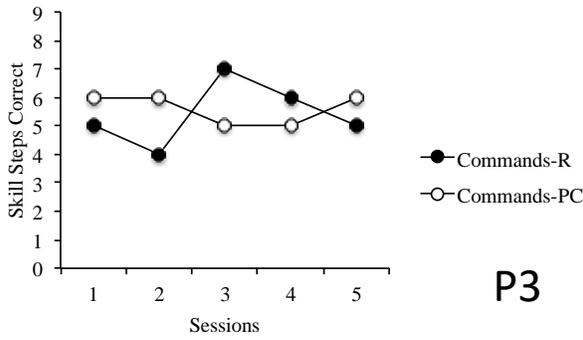
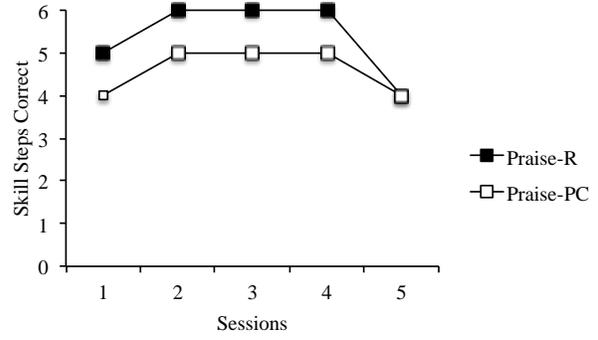
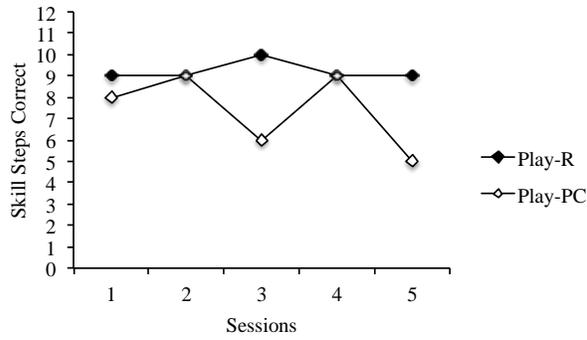
Table 1. Task analysis steps for four specific skills

<p style="text-align: center;">How to Play with Your Child (12 steps)</p> <ul style="list-style-type: none"> • Follows the child’s lead. • Plays with toys and makes statements • Allows for repetitive play • Does not criticize or correct • Verbally praises at least 2 times • Gives praise immediately (within 5 s.) • Plays at child’s level • Describes child’s actions • Describes child’s social behaviors • Identifies/names child emotions • Allows child to attempt first before assist • Doesn’t ask more than 3 questions 	<p style="text-align: center;">Using Effective Commands (9 steps)</p> <ul style="list-style-type: none"> • Command is specific • Command is direct • Does not use chain commands • Does not repeat command more than 3X • Does not begin command with “Let’s” • Does not form command as question • Command is positive • Uses warnings to signal transitions • Delivers reinforcement
<p style="text-align: center;">Positive Attention and Praise (6 steps)</p> <ul style="list-style-type: none"> • Praise contingent on approp. behavior • Specific Praise • Models self-praise • Gives praise immediately (within 5 s.) • Does not pair praise with put-down • Paris praise with smile or eye-contact 	<p style="text-align: center;">Ignoring (5 steps)</p> <ul style="list-style-type: none"> • Ignored for at least 10 s. • Avoids eye contact during child behavior • Avoid verbal exchange with child • Avoids physical contact with child • Provides attention within 5 s of appropriate behavior

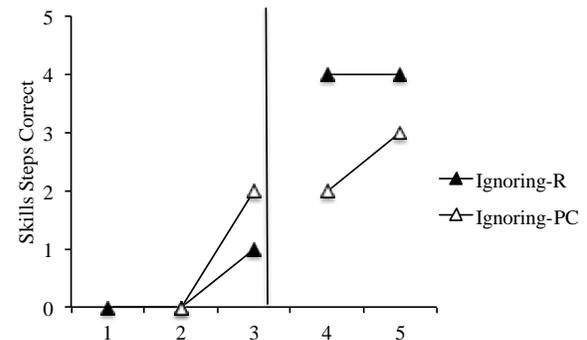
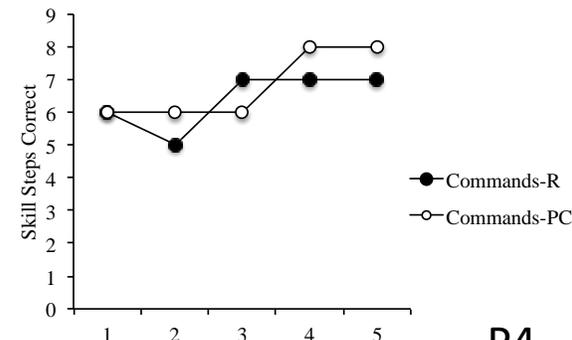
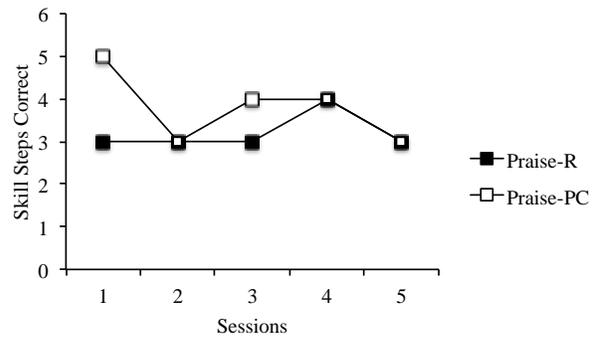
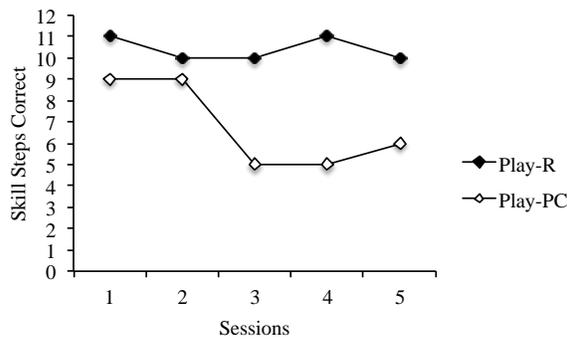
Note: more specific explanations and clarifications were included on the tracking sheet, but are simplified on this table for ease of reporting.

Figure 1. A comparison of scores from the parent-child interactions (PC) and the role-play (R) observations across all 5 sessions for participants P1-P4.





P3



P4

Note: R = Role-Play; PC = Parent-Child Interaction

Table 2. Accuracy, Similarity and Reliability of Role-Play and Parent-Child Observations

	Accuracy		Similarity	Reliability	
	R	PC	R & PC	PC	R
P1	41%	52%	71%	69%	64%
P2	55%	47%	39%	57%	71%
P3	69%	55%	72%	79%	86%
P4	55%	52%	71%	38%	100%
Play	74%	59%	40%	63%	88%
Praise	63%	61%	75%	56%	81%
Commands	57%	69%	61%	79%	71%
Ignoring	25%	18%	69%	33%	80%
Total	55%	52%	63%	59%	80%

Note: R = Role-Play; PC = Parent-Child Interaction; Accuracy = percent of steps correct;
 Similarity = percent of steps within one step between Role-play and Parent-Child Interaction;
 Reliability = percent of steps within one step within Role-play or Parent-Child Interaction