

**“Why Don’t You Act Like This at Home?!” Parent and Child Reactivity During In-Home Dyadic Parent-Child Interaction Coding System (DPICS) Coded Observations**

by

Timothy Thornberry, Jr., MS

A dissertation submitted to the Graduate Faculty of  
Auburn University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Auburn, Alabama  
August 3, 2013

Keywords: behavior observation, reactivity, assessment, parent-child interaction

Copyright 2013 by Timothy Scott Thornberry, Jr., MS

Approved by

Elizabeth Brestan-Knight, Chair, Associate Professor of Psychology  
Richard Mattson, Associate Professor of Psychology  
Martha Escobar, Associate Professor of Psychology  
Frank Weathers, Professor of Psychology  
John Dagley, Associate Professor of Counseling Psychology

## Abstract

The field of clinical child psychology has campaigned for evidence-based practice, with specific initiatives including the dissemination of evidence-based treatments and the development of evidence-based assessment guidelines. More work is needed in expanding the empirical literature regarding evidence-based assessment. This is especially true of analog behavior observations (ABOs). A specific threat to the external validity of ABOs is the potential reactivity participants may experience. This study sought to support the external validity of the Dyadic Parent-Child Interaction Coding System (DPICS), an ABO used to measure parent-child interactions, and pilot a parent-report measure of parent and child reactivity. Twenty-seven parent-child dyads participated in DPICS observations either in the home or clinic setting, and parents completed a new measure of parent and child reactivity: the TORQ. Results showed that behavioral differences did occur for both parents and children across observation settings, but these differences were not entirely accounted for by reactivity. However, TORQ scales did predict significant amounts of variance in DPICS composite scores of parent and child prosocial behavior and child compliance during various segments of the DPICS. Limitations, implications, and future directions in research are discussed. This study highlights the importance of considering reactivity when gathering observational data and offers a potential solution for documenting this reactivity.

## Acknowledgements

This project is the culmination of contributions from a variety of people that deserve recognition and, at least, money or physical labor by the primary author as compensation for their time and effort throughout this project. First, my deepest gratitude to the members of the Parent-Child Research Lab who assisted in data collection, DPICS coding, or who endured lab meetings or practice proposal run-throughs related to this project. Next, my great thanks for an excellent mentor, Dr. Elizabeth Brestan-Knight, for her patience, support, and guidance throughout my graduate career – it goes without saying that I would not be the professional I am today without her wisdom. I would also like to thank my committee members for their valuable input and time spent reviewing and contributing to this project. Finally, I must thank my family and friends for their unending support and love that helped me endure college, graduate school, and internship. Without the consistent social praise and contingent reinforcement of my parents, grandparents, and, especially, my wife, I would not have become the first member of my family to earn a doctorate. Also, my deepest apologies to them that it was not a medical doctorate.

## Table of Contents

Abstract .....	ii
Acknowledgments .....	iii
List of Tables .....	vi
List of Abbreviations .....	vii
Introduction.....	1
Evidence-Based Assessment.....	1
Barriers to EBA.....	3
EBA of Children and Families.....	7
Pros and Cons of Behavior Observation.....	9
A Continuum of Behavior Observations.....	11
Behavior Observations and Child Conduct Problems .....	13
Reactivity .....	15
In-Home Observations .....	21
Parent-Child Interaction Therapy .....	26
The Dyadic Parent-Child Interaction Coding System (DPICS) .....	27
Psychometrics of the DPICS.....	28
Dissemination of PCIT .....	29
Study Goals.....	30
Hypotheses .....	30

Method .....	33
Participants .....	33
Measures .....	35
Procedure .....	38
Results .....	41
Discussion .....	52
References .....	68
Appendix .....	81

## List of Tables

Parent and Child Behaviors and Respective DPICS-III Codes .....	90
DPICS-III Parent and Child Composite Categories and Respective Formulae .....	91
Demographic and Behavior Rating Information by Observation Setting .....	92
BASC-2 and ECBI Scores by Observation Setting .....	96
TORQ Descriptive Statistics.....	97
TORQ Correlations between Scales .....	98
TORQ Planned Comparisons between Observation Settings.....	99
DPICS Interrater Reliability by Observation Setting.....	100
DPICS Means and Standard Deviations during 5-Minute CLP by Observation Setting.....	102
DPICS Means and Standard Deviations during 5-Minute PLP by Observation Setting .....	104
DPICS Means and Standard Deviations during 5-Minute CU by Observation Setting .....	106

## List of Abbreviations

ABO	Analog Behavior Observation
APA	American Psychological Association
BASC	Behavior Assessment Scale for Children
CDI	Child-Directed Interaction
CLP	Child-Led Play
CT	Compliance Test
CU	Clean-Up
DPICS	Dyadic Parent-Child Interaction Coding System
EBA	Evidence-Based Assessment
EBI	Evidence-Based Instrument
EBM	Evidence-Based Medicine
EBPP	Evidence-Based Psychological Practice
EBT	Evidence-Based Treatment
ECBI	Eyberg Child Behavior Inventory
ICC	Intraclass Correlation
PCIT	Parent-Child Interaction Therapy
PDI	Parent-Directed Interaction
PLP	Parent-Led Play
PRIDE	Praise, Reflection, Imitation, Description, Enjoyment

TORQ Thornberry Observation Reactivity Questionnaire

WU Warm-Up



## **Introduction**

The recent evidence-based practice movement has been called the “most important trend in health care in the past two decades” (Hunsley & Mash, 2010, p. 3). Although evidence-based practice in clinical psychology has been developing for quite some time, a model was officially adopted by the American Psychological Association (APA) in 2006 when the APA Presidential Task Force on Evidence-Based Practice defined evidence-based psychological practice (EBPP) as “the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences” in order to “promote effective psychological practice and enhance public health by applying empirically supported principles of psychological assessment, case formulation, therapeutic relationship, and intervention” (p. 273).

### **Evidence-Based Assessment**

Despite the advancements in developing EBPP guidelines, many researchers argue that this progress is incomplete, in part, due to an overemphasis on establishing evidence-based treatment (EBT) at the relative neglect of developing evidence-based assessment (EBA; Achenbach, 2005; Hunsley & Mash, 2010, 2011; Jensen-Doss, 2011). This imbalance may be partially explained by fears that a lack of evidence supporting psychotherapeutic intervention places practitioners at risk of becoming obsolete in the face of the evidence-based medicine (EBM) movement. Why would clients need therapy of questionable utility when documented, effective medications are available? In contrast, the task of psychological assessment, “a unique and defining feature of the profession,” (Hunsley & Mash, 2011, p. 76) is not at risk of becoming

obsolete as a result of the EBM movement; therefore, the documentation of its empirical support may not be as high a priority to the field. Furthermore, an extensive history in psychology related to assessment may make “psychologists complacent about the nature and value of psychological assessment methods and practices” (Hunsley & Mash, 2011, p. 77). Thus, we develop “a steadfast belief in the intrinsic worth” (Mash & Hunsley, 2005, p. 363) of these assessments that is unfounded, leading to their continued use despite insufficient empirical grounding and insufficient proof that assessment improves clinical outcomes.

Whatever the reason for this disparity in research between EBT and EBA, the neglect of EBA could undermine the entire EBPP enterprise. As Hunsley and Mash (2011) point out, “the evaluation of, and, ultimately, the identification of EBTs rests entirely on the assessment data, [and] ignoring the quality of psychological assessment instruments and the manner in which they are used places the promotion of evidence-based psychological practice in jeopardy” (p. 82). Jensen-Doss (2011) expands on this concern and offers several reasons why researchers should focus on EBA. First of all, given that assessment is used to convey research and clinical findings, failure to use EBAs could disrupt communication between clinicians and researchers, making it difficult to integrate research and practice. Second, given that EBTs are typically organized by diagnosis and that these diagnoses are generated by assessment, failure to use EBAs could result in selecting an inappropriate EBT to implement with a given client or failing to collect all relevant information needed to use an appropriate EBT effectively (Hunsley & Mash, 2010; Jensen-Doss, 2011).

The abovementioned concerns thus justify the need for the field to develop EBA guidelines in order to ensure the integrity of EBPP. Progress has been made and can be credited, in large part, to the work of Eric Mash and John Hunsley, who spearheaded special journal

sections in 2005 in the *Journal of Clinical Child and Adolescent Psychology* (Mash & Hunsley, 2005) and *Psychological Assessment* (Hunsley & Mash, 2005), have written numerous book chapters, and edited a book dedicated to furthering the EBA movement (Hunsley & Mash, 2008a, 2008b, 2010, 2011). These authors define EBA as “an approach to clinical evaluation that uses research and theory to guide the selection of constructs to be assessed for a specific assessment purpose, the methods and measures to be used in the assessment, and the manner in which the assessment process unfolds” (Hunsley & Mash, 2011, p. 77). Thus, EBA is not merely the administration of evidence-based instruments (EBIs), but “a decision-making task in which the psychologist must iteratively formulate and test hypotheses by integrating data that may be incomplete or inconsistent” (p. 77). Thus, ideally, EBA guidelines would assist clinicians in selecting a battery of EBIs for a given client, administering these instruments in a context supported by scientific evidence (i.e., a given client population, assessment purpose, etc.), and interpreting possibly discrepant data in an evidence-based manner (Jensen-Doss, 2011).

### **Barriers to EBA**

Unfortunately, Hunsley and Mash’s EBA conceptualization is an unattained ideal at present as there are several barriers that impede the development of EBA guidelines. Some of these barriers reflect those encountered by the EBT movement; namely, there is disagreement regarding how much and what kind of evidence is needed to attain “evidence-based” status. Similar to the EBT movement, there have been several attempts to establish criteria for EBIs and to list measures that meet those criteria (Hunsley & Mash, 2008b, 2011); however, there is only some overlap between these lists, suggesting the existence of biases and preferences for certain assessment modalities over others (e.g., rating scales over observations). We cannot hope to reach a consensus on EBA processes and decision-making aids if we cannot agree on what

individual measures can be included in those processes. Thus, the EBA movement seems to have stalled, with the current focus of the field on proposing criteria for EBIs.

Within clinical psychology, only the Society of Pediatric Psychology has attempted to develop EBA guidelines (Cohen, La Greca, Blount, Kazak, Holmbeck, & Lemanek, 2008). Hunsley and Mash (2011) offer that there may be a reluctance to push for assessment guidelines beyond existing, generic standards (e.g., *The Standards for Educational and Psychological Testing*, American Educational Research Association, APA, & National Council on Measurement in Education, 1999) that offer “no guidance on the level of psychometric adequacy an instrument should have” (p. 83). This reluctance is alluded to in the EBPP Task Force report (2006): “APA also recognized the risk that guidelines might be used inappropriately by commercial health care organizations not intimately familiar with the scientific basis of practice to dictate specific forms of treatment and restrict patient access to care” (p. 271). Thus, not only is there disagreement amongst psychology professionals about what constitutes EBA, but there is also a concern that those outside of psychology may misuse guidelines at the expense of client care. These difficulties have slowed the development of EBA guidelines.

There is also a realization that EBTs and EBAs are typically diagnosis-specific despite the reality that most clients present with comorbid difficulties. Thus, it is uncertain how exactly clinicians are to apply EBTs and EBAs to clients presenting with multiple psychological concerns. That being said, outside of clinical psychology, there have been attempts to develop assessment guidelines for specific disorders by the American Academy of Child and Adolescent Psychiatry and the American Academy of Pediatrics (Hunsley & Mash, 2011). Hunsley and Mash (2010) argue that such a diagnosis-specific approach to EBA is required but remind clinicians that comorbidity is often the rule rather than the exception.

In addition to these difficulties shared by the EBT and EBA movements, EBA guidelines are further hindered due to a lack of psychometric “gold standard” analogous to the randomized control trial, often considered the pinnacle methodology in defining EBT (Jensen-Doss, 2011). Indeed, there are a multitude of psychometric characteristics to consider when selecting a measure, including the quality of supporting standardization data, a variety of reliability and validity indices, and clinical utility. As stated by Hunsley and Mash (2011), “assessment scholars, psychometricians, and test developers have typically been reluctant to set the minimum psychometric criteria necessary for specifying when an instrument is scientifically sound” (p. 83), leaving the decision of whether a given measure is “good enough” to individual clinicians.

Even if EBA guidelines could be agreed upon and established, there are anticipated barriers that would hinder their implementation (Hunsley & Mash, 2011; Jensen-Doss, 2011). As Hunsley and Mash (2011) point out, even if needed EBA guidelines are established, it does not necessarily mean that clinicians will follow the guidelines. For example, Jensen-Doss (2011) reports that, even when empirical support is available to guide clinical work, clinicians do not always adhere to such recommendations. For example, clinicians often utilize unstructured clinical interviews instead of their structured counterparts despite evidence questioning the validity of the former (Basco et al., 2000; Rettew, Lynch, Achenbach, Dumenci, & Ivanova, 2009). Similarly, Hunsley and Mash (2011) argue that “the weight of scientific evidence is not being used to its fullest extent in a number of areas in the domain of psychological assessment” (p. 78). These authors support this statement by pointing out survey data that show widespread use of assessments with insufficient research support (Hunsley & Mash, 2010, 2011).

Why do clinicians’ practices veer from research recommendations? According to Jensen-Doss (2011), surveys indicate that clinicians’ decision making is largely guided by practical

concerns, which was found to be the largest and only independent predictor of standardized assessment use (Jensen-Doss & Hawley, 2010). Indeed, a survey by Luebbe, Radcliffe, Callands, Green, and Thorn (2007) found that graduate students in clinical psychology programs reported that “the nature of a treatment’s empirical support was among the least important factors influencing students’ treatment planning decisions” (as cited in Hunsley & Mash, 2010, p. 5). Rather, professional psychologists seem to look to colleagues and the ease with which a treatment can be learned when deciding which treatments to utilize (Nelson & Steele, 2008). Thus, a potential key to ensuring the implementation of EBA guidelines, when they are developed, is to document their clinical utility, a typically understudied area of assessment (Hunsley & Mash, 2011). Characteristics of measures tied to clinical utility can include available languages, training required for administration and interpretation, administration and scoring time needed, computerized scoring options, and cost (Jensen-Doss, 2011).

Another possible EBA-hindering factor relates to recent cost-containment strategies brought about by managed care. According to Cantor and Fuentes (2008), the rise of managed care during the past few decades has discouraged psychological assessment practices by decreasing the number of testing referrals and the reimbursement fees provided for assessment. This cost-cutting environment has made it more difficult for psychologists to justify providing comprehensive assessment services. Indeed, a survey by Piotrowski, Belter, & Keller (1998) found that clinicians were providing less testing with fewer instruments, including instruments ranked as most important by those surveyed (e.g., Rorschach, Thematic Apperception Test, Minnesota Multiphasic Personality Inventory, Wechsler scales). Furthermore, clinicians are unlikely to be reimbursed for time spent scoring, interpreting, and writing assessment reports (Eisman et al., 2000; Stout & Cook, 1999; Turchik, Karpenko, Hammers, & McNamara, 2007).

Given this impact of managed care on some of the most esteemed assessment instruments in our field, it is likely that efforts to develop EBA in common practice will face an uphill struggle.

In sum, there are numerous barriers to the development of EBA, including: insufficient research motivation relative to EBT; complacency amongst psychologists that our current assessments are “good enough” and our assessment role is safe from other fields; disagreement about what constitutes “evidence based;” absence of a psychometric “gold standard” with which to determine a measure’s quality; and practical, cost, and clinical utility considerations that largely guide a clinician’s treatment and assessment selection contrary to empirically-determined, best-practice guidelines. However, progress in EBA is being made and more is needed to ensure the best quality assessment and treatment services are rendered to clients. It seems current research efforts need to not only form an empirical foundation and consensus for EBA, but also achieve relevance (i.e., external and predictive validity) and clinical utility for practitioners if EBA is to become common practice.

### **EBA of Children and Families**

Similarly to the aforementioned general state of EBPP, the field of child clinical psychology currently faces the need to develop EBA practices to accompany the recent push for EBTs (Chambless et al., 1996; Chambless et al., 1998; Mash & Hunsley, 2005). Mash and Hunsley (2005) state that many child and family assessments are used despite inadequate empirical support. These authors propose that EBAs of children and families should have the traditionally-emphasized promise of adequate reliability and validity as well as strong evidence of clinical utility. Along with psychometric rigor, researchers have argued that an efficient approach to child and family assessment utilizes a multi-stage process, beginning broadly and narrowing in focus as data accumulate (Achenbach, 2005; McMahon & Frick, 2005). Thus, an

EBA approach for children may utilize a two-staged process that begins with broadband, screening assessments to confirm clinically-relevant levels of symptomatology suggested by the presenting issue as well as identify possible comorbid issues. This broadband wave of assessment would then be followed with a narrowband, focused assessment battery of instruments with sufficient specificity to the identified problems of interest during the first wave of assessment. Such an approach to EBA of children would ensure comprehensive yet efficient services.

Furthermore, given the unique influences of family, peers, and developmental changes on children, researchers (Achenbach, 2005, Mash & Hunsley, 2005) advise that, relative to adult assessments, comprehensive assessments of children include a larger number of assessments as well as a multimodal approach, which may incorporate input from multiple informants (e.g., parents, teachers, and peers) and behavioral observations in multiple contexts (e.g., clinic, school, and home). Despite this proposal that EBA of children include behavioral observations, many practitioners neglect to use such observations consistently in comprehensive assessment practices (e.g., Palmiter, 2004). The underuse of observations does not appear to be indicative of incompetence, a lack of awareness, or disagreement as to the importance of behavioral observations in child assessment. For example, Cashel's (2002) survey of practicing psychologists indicates that clinicians hold behavior observations to be one of the most important assessment procedures second to the clinical interview. Further, Cashel argues that the use of structured observations has become increasingly prevalent in clinical practice.

However, a survey by Palmiter (2004) suggests otherwise. This survey of 309 clinicians who served children and adolescents revealed that less than half (41.7%) actually used naturalistic observations in their assessment practice and 59% reported that they would prefer to



use naturalistic observations if the decision was only based on their professional judgment and the preferences of their clients. These percentages of actual and preferred use are much lower than those associated with family interviews (89.5% and 95.6%, respectively), individual child/teen interviews (83.1% and 90.8%, respectively), and previous treatment record review (70.2% and 88.1%, respectively). Palmiter summarizes that “there appears to be a gap between what research suggests is best practice and what actually occurs in the clinical trenches” (p.126).

There are many potential reasons for this research-practice gap in child assessment methods. Palmiter’s (2004) survey found that the top three factors clinicians identified as influencing their choice of assessments were ethical concerns, organizational pressures, and theoretical orientation. Unfortunately for the EBA movement, one fifth of Palmiter’s sample indicated that research findings were unimportant in clinicians’ assessment decision process. This is interesting given that ethical practice requires the use of interventions and assessments that are backed by adequate research to ensure that harm is not done. Thus, in order for EBA guidelines for child and family assessment to take hold in practical settings, recommended assessment modalities, such as behavior observations, need to display incremental utility at an acceptable cost to the clinician (or other relevant third-party).

### **Pros and Cons of Behavior Observation**

Direct observation of child and family behavior offers data to the clinician that are not available through more efficient assessment modalities such as clinical interviews and paper-and-pencil measures. According to Gardner (2000), direct observation allows the clinician to witness behavioral processes in detail that “would be very hard for participants to access through self-report, as much of the behavior seen during encounters of interest (e.g., family conflict) may be automatic and fast moving” (p. 187). Furthermore, Aspland and Gardner (2003) report that

observations are “invaluable for planning interventions and evaluating outcomes” (p. 136). Specifically, these authors argue that observations eliminate a potential semantic problem in that behaviors of interest are defined by the researcher rather than the parents, who may report on their child’s behavior using idiosyncratic definitions that may be systematically biased (e.g., parent moods or expectations of treatment; Eddy, Dishion, & Stoolmiller, 1998; Fergusson, Lynskey, & Horwood, 1993; Patterson, 1982; Prescott et al., 2000; Richters, 1992). These biases may have less effect on observational data compared to self-report data (Aspland & Gardner, 2003). Not only may observational data protect against systematic bias due to expectancy effects, but they may also be more sensitive to parent and child behavioral change following treatment (Aspland & Gardner, 2003; Forgatch & DeGarmo, 1999; Patterson, 1982; Sanders, Markie-Dadds, Tully, & Bor, 2000; Webster-Stratton, 1994, 1998) and may better predict long-term outcomes (Patterson & Forgatch, 1995).

Despite these potential benefits of observational assessment, there are drawbacks to observational methods that impede their widespread implementation. One disadvantage relates to time costs associated with training observers, conducting observations, coding observations, and monitoring the reliability of observers (Frick & McMahon, 2008; Gardner, 2000; Margolin et al., 1998). Another concern raised by researchers relates to the lack of evidence that behavior observations generalize across relevant settings (Aspland & Gardner, 2003; Gardner, 2000; Hartmann & Wood, 1990; Kazdin, 1982; Pett, Wampold, Vaughan-Cole, & East, 1992). Often, behavior observations are conducted in settings (e.g., clinics and labs) that are different from families’ or children’s typical settings (e.g., home, supermarket, school) or include structured task demands unfamiliar to families (Gardner, 2000). As Gardner (2000) demonstrates, the generalizability of observational data to relevant settings is crucial if these data will inform the

behavioral goals of treatment. If observed behavior differs across settings, then our observational data may be irrelevant to the presenting problem. If such nonrepresentative observations form the basis of our treatments, then there is a chance that treatment gains will vary across settings or may not generalize at all outside of the clinical setting. Thus, more research needs to document the generalizability of observational data if observations are to be included in EBA of children and families.

### **A Continuum of Behavior Observations**

There are many forms of behavioral observation available to researchers and clinicians, each with its benefits and detriments. Broadly, behavioral observations can be viewed as existing along a continuum between purely naturalistic and highly structured formats, the latter known as analog behavioral observations (ABOs). Naturalistic observations place minimal constraints on participant behavior and minimize observer involvement in order to bolster the external validity of obtained data. Furthermore, naturalistic observation allows researchers to study certain behaviors that may be unethical or impossible to recreate in a laboratory or clinic setting (Pepler & Craig, 1995). However, when conducting naturalistic observations, it is uncertain whether the behavior(s) of interest will manifest during the observation, and, given the high cost associated with observation, many researchers may not be willing to take such a chance.

In contrast, ABOs seek to create some control over the observation setting in order to “derive valid and cost-effective estimates of a client’s behavior, thoughts and cognitive processes, emotions, and physiological functioning, and of interactions between the client and others” (Haynes, 2001, p. 73). The amount of control can vary greatly from highly contrived tasks, situations, and settings to relatively natural tasks in contrived settings to naturalistic situations with minimal restrictions (Heyman & Slep, 2004). Although such control arguably

threatens the external validity of gathered data by placing participants in unnatural settings (i.e., the lab) and asking them to perform certain behaviors at certain times, structured ABO can increase “the consistency of sampling between individuals or across time,” and “the likelihood of certain behaviors arising, which then allows comparisons to be made” in an efficient manner (Aspland & Gardner, 2003, p. 137; Gardner, 2000; Heyman & Slep, 2004). ABO with structured tasks can also improve reliability by “decreasing the range of possible situational influences on the behavior” of interest (Gardner, 2000). ABOs may also provide sufficient experimental control to minimize the need for inference and allow researchers to observe underlying processes that may be unobservable in uncontrolled settings; this added control can also allow researchers to test hypotheses about specific causal relations and facilitate functional analyses of behavior while controlling for unwanted sources of variance (Aspland & Gardner, 2003; Heyman & Slep, 2004; Margolin et al., 1998). By providing a standardized context in which to observe behavior at various times, ABOs can also be used to reliably assess behavioral change in treatment outcome studies (Heyman & Slep, 2004). In sum, as compared to naturalistic observations, ABO allows for more efficiency and cost-effectiveness while improving the reliability and internal validity of gathered data at the possible expense of generalizability of findings to other settings.

ABOs are particularly well-suited to provide the objective clinical assessment data needed to build empirical support for various child-focused interventions (Mori & Armendariz, 2001). As with observational data in general, the benefits of using ABOs in the context of child and family treatment include strong clinical utility via incremental utility (i.e., ability to provide data beyond what is provided by self-report measures), treatment utility (i.e., ability to incorporate data into treatment), and their sensitivity to change brought about by treatment (Haynes, 2001; Heyman & Slep, 2004).

Despite the positive aspects of using ABOs in the treatment of child clinical populations, as with behavior observations in general, there are limitations to using this form of assessment that inhibit the adoption of ABO as a preferred assessment tool in clinical practice (Mash & Foster, 2001). Cost barriers and accessibility concerns hinder the dissemination of most ABOs into clinical settings (Brestan-Knight & Salamone, 2011; Heyman & Slep, 2004; Mash & Foster, 2001). In addition, many child-focused analog measures lack reliability and validity data, evidence of standardization, and ecological and construct validity (Haynes, 2001; Heyman & Slep, 2004; Mash & Foster, 2001; Mori & Armendariz, 2001; Roberts, 2001). Given this lack of accessibility and empirical support, many clinicians make adaptations of existing ABOs without empirical justification in order to make them more clinic-friendly, further jeopardizing the reliability and validity of ABOs (Heyman & Slep, 2004). Thus, more research is needed to bolster the reliability, validity, and utility of many ABOs in order to justify their use in the evaluation and treatment of child psychopathology (Haynes, 2001).

### **Behavior Observations and Child Conduct Problems**

The full continuum of behavior observations has been used to assess child conduct problems in a variety of settings and for a variety of purposes (e.g., diagnosis, case conceptualization, treatment planning; Frick & McMahon, 2008). Generally, these observations can help identify functional relationships between the child's disruptive behavior and the behaviors of those in the child's environment, which can inform behaviorally-based interventions (Frick & McMahon, 2008). As mentioned previously, such data may be inaccessible via rating scales and interviews due to bias or unawareness on the part of the informant. When conducted within the context of behavioral parent-training programs, ABOs are used to collect behavioral data regarding the functioning of the parent-child dyad, specifically for commonly-used

parenting practices (e.g., use of commands and praise) known to be critical factors in the development of child conduct problems (Brestan-Knight & Salamone, 2011; Frick & McMahon, 2008; Gardner, 2000; Haynes, 2001; Heyman & Slep, 2004; McMahon & Frick, 2005; Mash & Foster, 2001; Roberts, 2001). ABOs are also used to monitor treatment progress and outcomes during family-based behavioral treatments of child conduct problems (Frick & McMahon, 2008).

Regarding the status of the literature base supporting ABOs of parent-child interactions, much more research is needed (Brestan-Knight & Salamone, 2011; Roberts, 2001). Roberts' (2001) review of ABOs concluded that analogs of free play, parent-directed play, and parent-directed chores are all psychometrically underdeveloped in the domains of test-retest reliability, clinical utility, and normative data. Along with these psychometric concerns, researchers acknowledge a lack of data regarding the optimal number and duration of observations needed during treatment and the added expense associated with using these behavior observations of parents and children as an assessment component (Mash & Foster, 2001; McMahon & Frick, 2005). Also, given the dynamic nature of validity (i.e., psychometric properties of assessments must be shown to generalize to new settings, with new populations, and for new purposes; Haynes, 2001; Heyman & Slep, 2004), ABOs must have documented research support in order to be used in different settings or for different purposes. Furthermore, Frick and McMahon's (2008) recent assessment of several well-known ABOs used to assess child conduct problems notes a continued need to establish normative data. Thus, norm data and more reliability, validity, and clinical utility evidence are needed to justify the use of ABOs in clinical research and practice.

## **Reactivity**

An as-yet unmentioned criticism of ABOs related to generalizability is the concern that the observation participants may behave differently during the observation relative to how they may behave in the naturalistic setting (e.g., in the home, at school). This process is known as “reactivity” or “observer effect” (Aspland & Gardner, 2003). As mentioned in Kazdin’s (1982) comprehensive article, researchers have considered reactivity effects on observational data since at least the 1940’s, but, the effects of reactivity are equivocal. Kazdin explains that although obtrusive observation can cause a participant to behave differently, this is not always the case. He then reviews four different paradigms that have been used to examine reactivity: observer-present, instructional-set variation, self-monitoring, and observer-reaction paradigms. All of these paradigms produce mixed results, making it difficult to offer a definitive conclusion as to the effects of reactivity on observational data. Still, several studies show that reactivity should at least be considered by clinicians and researchers, as it has the potential to produce significant changes in behavioral data.

For example, Kazdin (1982) cites a study by White (1977) that, using the observer-present paradigm, found a decrease in family activity during the observer’s presence. In contrast, using the instructional-set variation paradigm, Zegjib and Forehand (1978) found that mothers became more active (i.e., played more, gave more commands) when they were told they were being observed. Reactive effects can also be produced using self-monitoring procedures. Herbert and Baer (1972) report as much when they asked mothers to count the number of times they attended to their children’s positive behaviors; these instructions were found to increase maternal attention and improve child behavior. A final, noteworthy example of reactivity that moves beyond the scope of this project is found in the observer-reaction paradigm mentioned by Kazdin

(1982). This paradigm investigates how observers' behaviors can change as a function of observing others' behavior. This becomes relevant in the coding of family behavior using ABOs in that observing certain family behaviors may in turn affect coders' coding behavior. This is an interesting concept that should be studied in future studies but is beyond the scope of this project.

After reviewing these various paradigms for studying reactivity, Kazdin (1982) concludes that the inconsistent results found may indicate that reactivity has weak effects. If reactivity produces weak effects, then perhaps it is a non-issue compared to the large effects of some behavioral interventions. However, it is also possible that reactivity is dependent on other variables, such as the valence of the behavior (Kazdin, 1982). In other words, reactivity may increase positively-valenced behaviors (e.g., praise) and decrease negatively-valenced behaviors (e.g., criticism). Thus, reactivity may be a helpful tool in the context of a behavioral parent training program in that parents will be motivated to increase positive parenting behaviors while decreasing negative behaviors during obtrusive behavior observations. This unique evaluative context may give parents the practice they need to change their behavior, and thus their child's behavior, in a therapeutic direction. However, this is only true if the behavioral changes during obtrusive assessment generalize to the naturalistic setting. Also, the magnitude of reactivity effects may differ based on behavior valence. For example, Baum, Forehand, and Zegiob (1979) found that observer effects were greater on positively valenced behaviors like praise and playing with the child than they were on negatively valenced behaviors such as criticism. In other words, reactivity may motivate parents to "fake good" but parents may not be as effective at decreasing negative behaviors as they are at increasing positive behaviors.

Kazdin (1982) also explores theoretical explanations for reactivity. First, reactivity may be attributed to evaluation apprehension and socially-desirable responding. That is, people will



behave in such a way as to make the observer like them. However, as Kazdin points out, reactivity appears to have a larger effect on verbal rather than nonverbal behavior; thus, “people are more likely to monitor and control what they say than they are likely to monitor and control what they do” (Kazdin, 1982, p. 10). This may help account for the mixed data documenting the effects of reactivity as verbal data may indicate a larger effect than nonverbal behavior. During observations of parents and children, perhaps parents will speak differently to their child but may not be able to alter nonverbal behavior as effectively.

Stimulus control may also help explain reactivity effects. Namely, an obtrusive assessment situation can exert stimulus control over certain participant behaviors depending on the participant’s interpretation of the purpose of the assessment (Kazdin, 1982). This may help explain habituation effects on reactivity, i.e., reactivity effects are believed to dissipate over time as the obtrusiveness of the observation decreases and the situation loses its stimulus control over the participant’s behavior. In other words, parents learn that they can “relax” in subsequent observations and behave as they typically do outside of the clinic.

Finally, feedback and self-regulation may help explain reactivity in that obtrusive observation may make certain behaviors more salient to the observed – they pay more attention to certain behaviors they believe relevant to the observation situation/purpose (Kazdin, 1982). By paying attention to these behaviors, participants can give themselves feedback, reinforce or punish themselves, and regulate those behaviors in the future. This helps explain reactivity effects brought about through self-monitoring of behavior.

Kazdin (1982) warns that these three theoretical explanations for reactivity are not an exhaustive list and are post hoc. Further, he argues that research on reactivity has not been guided by theory and data inconsistently support the notion of reactivity. He goes on to warn that

researchers need to better understand reactivity given its threats to the external and internal validity of observational assessment. This 30-year-old recommendation is still relevant given the recent developments in EBPP.

More recent research related to observation reactivity on parent and child behavior has been conducted by Gardner and colleagues (Aspland & Gardner, 2003; Gardner, 2000). Gardner's (2000) recent review echoes Kazdin's (1982) assessment of the equivocal nature of data supporting reactivity. Furthermore, Gardner presents two additional methods used to assess reactivity. The first method compares parent and child behaviors observed under varying levels of observer intrusiveness (e.g., audiotape versus audiotape and observer present). However, according to Gardner (2000), this method has failed to find differences in parent and child negative behavior or parent commands in multiple studies (e.g., Bernal, Gibson, William, & Pesses, 1971; Jacob, Tennenbaum, Seilhamer, & Bargiel, 1994; Johnson & Bolstad, 1975), suggesting that observer intrusiveness may not alter parent or child behavior.

Researchers have also tried to study reactivity by comparing participant behavior over time to determine if participants' habituation to the observation setting leads to decreases in reactivity effects. Based on this model, behavior observed during initial session segments or initial sessions in a series of observations should be atypical when compared to subsequent observation sessions due to stronger reactivity effects (Gardner, 2000). This method has also produced mixed results. For example, Hughes, Carmichael, Pinkerton, and Tizard (1979) found that, during four separate home-based observations of preschoolers and their mothers, there were no conversational differences across sessions with the exception of children talking more to observers during the first session; however, there were no differences in mother or child verbal behavior. Similar results were found by Johnson and Bolstad (1975), who found no differences

between observation sessions or within observation sessions (i.e., first 15 min versus subsequent 15-min segments).

In contrast, Kier (1996) found during preschool siblings' interactions that siblings showed more independent play during the first 10 min of an hour-long observation and were more proximal to their sibling as compared to the last 10 min of the session. Kier suggests that this indicates a habituation effect on reactivity. Unfortunately, as pointed out in Gardner's (2000) review, it is unknown which 10-min portion of Kier's (1996) observation was more representative of the entire segment; instead, Kier assumes that reactivity negatively affected the first segment. Even if Kier's study provides evidence of reactivity, it is unknown if such effects observed in siblings will also occur in parent-child interactions. Overall, these studies illustrate that, despite evolving research methodology, there is still uncertainty regarding the effects of reactivity on child and parent behavior.

Despite this uncertainty about the effects of reactivity, researchers have identified numerous factors believed to affect reactivity. Aspland and Gardner (2003) report that these factors may include: the conspicuousness of the observation process; whether participants have the opportunity to habituate to the observer's presence; the participants' understanding of the purpose of the observation; the demands of the observation; and the setting of the observation. In addition, Gardner (2000) says that reactivity may also be affected by participant characteristics, such as age and gender of the child or parent. For example, studies suggest that fathers' behavior may be more affected by reactivity than mothers' behavior (Lewis et al., 1996; Russell, Russell, & Midwinter, 1992). Also, parent psychological functioning may play a role in reactivity. In a study by Johnson and Lobitz (1974), distressed parents had difficulty manipulating their children into behaving well during home-based observations. Thus, even in situations where parents may

be reacting to the presence of an observer, their attempts to change their own behavior or their child's behavior may be ineffective if they are distressed. It is also believed that reactivity has a smaller effect on young children (Aspland & Gardner, 2003). Thus, observers should not assume that all participants will be affected by reactivity equally.

Based on the aforementioned factors believed to affect reactivity, researchers have made recommendations for minimizing the effects of reactivity on observational data. Kazdin (1982) recommends that researchers and clinicians use unobtrusive measures if possible; however, doing so may not only hurt the psychometric integrity of the data (e.g., reliability and validity), but may also be unethical if participants are being observed unknowingly or are being deceived by confederates. Also, although naturalistic observation is unobtrusive, as mentioned earlier, such unstructured observations may not produce meaningful data if behaviors of interest do not spontaneously occur or if behaviors occur unreliably. Given the aforementioned cost concerns of conducting observations and the increasing focus of clinicians on managed care and practical concerns, naturalistic observations seem to be prohibitive.

Some researchers recommend that observers not code participants' behaviors initially, whether that be for the first 10 min of the observation or the entire first observation session (Dunn & Kendrick, 1980; Kier, 1996); others argue that no data should be excluded (Gardner, 2000). Other recommendations include allowing time for participants to become familiarized with the observation procedures, using the same observer across multiple observations, minimizing the number of observers present during the observation, avoiding interaction with the participants during recording, and minimizing the obtrusiveness of the observation equipment and procedures (Aspland & Gardner, 2003; Gardner, 2000; Kazdin, 1982). Finally, in line with

EBA recommendations, researchers suggest that the use of multiple assessment methods can help control for method variance caused by reactivity (Kazdin, 1982).

Many of these recommendations have been incorporated into existing ABO protocols. For example, the Dyadic Parent-Child Interaction Coding System (DPICS; Eyberg, Nelson, Duke, & Boggs, 2005) uses warm-up segments to allow for participant habituation to the observation situation and utilizes a bug-in-the-ear device and one-way mirror to minimize obtrusiveness. Ultimately, it is hoped that adhering to these recommendations will reduce reactivity effects and thereby maximize the generalizability of ABO data to more naturalistic settings such as the home or school.

### **In-Home Observations**

Given that observation setting is a proposed factor that may influence reactivity effects, and given the aforementioned concerns of the generalizability of observational data to real-world settings where behaviors of interest may occur, it may be productive to conduct ABOs in these real-world settings (i.e., the home). This generalizability concern may be especially pertinent when observing families of children with disruptive behavior. According to Aspland and Gardner (2003), “many researchers favor home observations in studies of parent-child interactions in families of conduct problem children, as they clearly provide a much closer approximation of the environment the parent and child normally interact in and are, therefore, likely to show greater validity” (p. 137; Jacob, Tennenbaum, Bargiel, & Seilhamer, 1995). Anecdotally, it seems plausible that children with behavioral problems may be less likely to exhibit misbehavior in an unfamiliar clinic setting in comparison with their more familiar home setting (however, see Masse & McNeil (2008) for a discussion of how in-home observations may also discourage misbehavior). Indeed, Johnson and Lobitz (1974) found that distressed parents

were unable to manipulate their children into behaving well during home observations. Unfortunately, there is a paucity of research comparing observational data gathered in a clinic setting with that collected in the home setting (Gardner, 2000; Rapoport & Benoit, 1975). Without these studies, it is ill-advised to assume that ABOs conducted in the clinic will be comparable to those conducted in the home setting, even if reactivity has a minimal effect (Gardner, 2000).

To illustrate, consider a study by Webster-Stratton (1985) that found moderate to high correlations between the behaviors of mothers and children observed during unstructured situations in the home and clinic. Specifically, mothers who were more directive in the clinic were also more directive in the home and children who were more compliant in one setting were also more compliant in the other. This would appear to support the validity of clinic-based observations. However, Webster-Stratton also observed that there were significant differences in the mean frequencies of these behaviors between settings. That is, mothers were more directive and used more praise in the clinic and children displayed more deviant behavior in the home. Thus, clinic-based, as compared to home-based, observations may yield data that approximate behaviors seen in the home, but these behaviors may differ in magnitude or intensity. It is uncertain what mechanisms produced these differences between settings. Perhaps reactivity played a role or perhaps other unknown factors associated with the home setting produced these differences. Indeed, it may be more difficult to control for extraneous variables and distractions when conducting observations in the home setting as opposed to the clinic or lab setting (Masse & McNeil, 2008). Also, Webster-Stratton employed an unstructured observation task; therefore, it is possible that these unstandardized observations simply produced unreliable data. Perhaps using more structured ABOs in a similar comparison between home and clinic settings will allow

sufficient experimental control to make clearer conclusions regarding reactivity and external validity of clinic-based observations.

There are a variety of ABO tasks that could be employed in the home setting, each of which may differ in their ability to pull for certain behaviors from parents and children and in how well they compare to various situations in real-world settings (i.e., their external validity; Gardner, 2000). Unfortunately, there are few studies that examine the differential representativeness of these tasks using video recording in the home. Although the studies reviewed so far would suggest that the added intrusiveness associated with video recording observations is unlikely to add to reactivity effects (see Bernal et al., 1971; Jacob et al., 1994; Johnson & Bolstad, 1975; Pett et al., 1992), Gardner (2000) warns that this conclusion cannot yet be made about home-based video recording given a lack of studies. Still, researchers suggest that observers familiarize families with video recording procedures prior to recording in order to minimize reactivity (Dunn & Kendrick, 1982; Gardner, 1987, 2000).

An unpublished longitudinal study by Gardner and colleagues (see Gardner, 2000) provides an exception to the lack of studies in this area. This study compared home-based observational data collected using structured and unstructured tasks. In order to minimize reactivity effects due to video recording, families were introduced to the recording procedures during a previous visit. Gardner and colleagues found significant correlations between structured and unstructured tasks, but also found that mother-child conflict observed during unstructured tasks was more-highly correlated with questionnaire and interview measures of conduct problems than was behavior observed during the structured task. These data suggest that more naturalistic, unstructured tasks may produce more valid (e.g., convergent validity) observational data in the home setting as opposed to structured tasks. However, the validity of these findings

may fluctuate depending on the focus of the observation (e.g., observing parent versus child behavior) and the stability of the observed behaviors over time. Standardization studies using ABOs are needed in order to make definitive conclusions about the validity of these observations. Furthermore, future studies can expand on Gardner's (2000) findings by comparing the relative predictive power of ABOs with varying levels of structure.

Although not conducted in the home setting, there is a recent, clinic-based study by Rhule, McMahon, and Vando (2009) using a nonreferred community sample of mothers and children that investigated parent-reported acceptability of various ABO tasks (child's game, parent's game, clean-up, and Compliance Test (CT; Roberts & Powers, 1988) and the representativeness of behavior captured by each task compared to typical interactions that occur in the home (e.g., how typical was each task in the family's home, how did the child's behavior during the task compare to typical home behavior, how did the parent's behavior during the task compare to typical behavior at home). Acceptability and representativeness were assessed using 7-point Likert-type scales.

This study found that mothers reported the lowest acceptability and representativeness with the highest-structured task: the CT. Indeed, 18% of CTs administered in this study had to be discontinued due to child distress or the parent or child requesting to stop the task whereas no other ABO task had to be discontinued. Furthermore, the child's game and clean-up tasks were rated as more acceptable and more representative than the parent's game. Overall, these results suggest that higher observation structure is perceived as less acceptable and less representative by parents.

Regarding the cause of these results, the authors hypothesize that parents' discomfort may have contributed to less acceptability of the CT. Specifically, parents reported feeling



uncomfortable inhibiting their behavior (e.g., praise, repeat command, discipline) during the CT. However, the authors did not directly assess parent comfort during each ABO task, rendering these parent reports anecdotal and qualitative in nature. Furthermore, the authors did not measure the parents' perceived difficulty with conducting the various tasks, which might have influenced their or their child's behavior such that more difficult tasks yielded less representative data or less acceptable reports. Future studies should document perceived difficulty and parent discomfort directly in order to isolate their influence on parent reports of acceptability and representativeness of various ABO tasks. A strength of this study includes the authors assessing acceptability and representativeness across two sessions separated by two weeks, allowing for the calculation of test-retest reliability for these measures. Although child and parent behavior was assessed during this study using an interval coding system, the authors do not comment on the stability of parent and child behavior within (i.e., across tasks within a session) or across observation sessions or comment on how variability in acceptability and representativeness may be attributable to variability in parent and child behavior.

Interestingly, Rhule and colleagues (2009) also found differences in acceptability and representativeness as a function of participant factors. Specifically, they found that mothers of girls reported the clean-up task to be less acceptable and less representative than did mothers of boys. Furthermore, they found that mothers of older children reported the parent's game to be less representative of typical home behavior than did mothers of younger children. These findings back the aforementioned recommendations to consider participant variables, such as child age and gender, when studying reactivity effects in ABO of child behavior problems.

## **Parent-Child Interaction Therapy**

Perhaps the utility of ABO is best illustrated by its use in an evidence-based treatment for children with disruptive behavioral disorders. Parent-Child Interaction Therapy (PCIT) is a parent-training program that seeks to improve the parent-child relationship and correct maladaptive parent and child behavioral patterns. PCIT is based on Hanf's (1969) two-stage model and is divided into two separate phases of treatment. The first phase is the Child-Directed Interaction (CDI) phase. During this phase, parents are taught the PRIDE skills, an acronym for the skills of Praise, Reflection, Imitation, Description, and Enjoyment; these skills are intended to improve the parent-child relationship by encouraging positive parental attention and selective ignoring of minor misbehaviors. Once parents reach behavioral criteria for using the PRIDE skills, therapy transitions to the second phase of treatment, known as the Parent-Directed Interaction (PDI) phase. This phase of treatment focuses on teaching the parent effective discipline techniques for correcting a child's disruptive or defiant behavior.

PCIT has been found to be effective in treating families of children with disruptive behavioral disorders (Nixon, Sweeney, Erickson, & Touyz, 2003; Schuhmann, Foote, Eyberg, Boggs, & Algina, 1998) as well as numerous other childhood psychological problems, including children with abuse histories (Chaffin et al., 2004; Ware, Fortson, & McNeil, 2003), separation anxiety disorder (Choate, Pincus, Eyberg, & Barlow, 2005), and mental retardation (Bagner & Eyberg, 2007). PCIT has also been shown to improve disruptive behaviors in children with comorbid medical problems such as diabetes (Miller & Eyberg, 1991) and cancer (Bagner, Fernandez, & Eyberg, 2004). Furthermore, the treatment effects of PCIT have been shown to generalize beyond the home to improve children's behavior in the school (Funderburk et al., 1998; McNeil, Eyberg, Eisenstadt, Newcomb, & Funderburk, 1991) and hospital (Bagner,

Fernandez, & Eyberg, 2004) contexts and to untreated siblings (Brestan, Eyberg, Boggs, & Algina, 1997).

PCIT differs from other parent training therapies in that treatment goals are accomplished through direct, *in vivo* instruction via a bug-in-ear device. This allows the therapist to coach the parent to use positive parenting skills (e.g., PRIDE skills, planned ignoring, consistent discipline) in real-time while providing immediate feedback regarding the parent's performance. This feedback, although at times informal and given throughout therapy sessions, is formalized using an ABO designed specifically for PCIT: the Dyadic Parent-Child Interaction Coding System (DPICS; Eyberg, Nelson, Duke, & Boggs, 2005).

### **The Dyadic Parent-Child Interaction Coding System (DPICS)**

The DPICS is used to code parent and child behaviors observed during a standardized, 25-minute play situation which is divided into three major segments: a Child-Led Play (CLP) segment, Parent-Led Play (PLP) segment, and a Clean-Up (CU) segment. During the first 10 minutes of observation, parents are instructed to allow the child to lead the play. The first 5 minutes of this segment is designated a "warm-up" (WU) segment that is intended to allow the dyad to acclimate to the play situation, minimize reactivity, and maximize the validity of observed behavior. The WU segment is not coded, however, the latter 5 minutes of the CLP segment are coded. The CLP segment is followed by a 10-minute PLP segment, during which the parent is instructed to lead the play. Again, the first 5 minutes of this segment is designated a WU segment and is not coded while the latter 5 minutes of PLP are coded. The final 5 minutes of observation consists of the CU segment, where parents are instructed to direct the children to clean up the toys by themselves. There is no WU segment for CU as it is rare that the child takes the full 10-minute period to successfully clean up the toys. From a separate observation room,

the therapist delivers instructions to parents via a bug-in-ear device so as to limit the obtrusiveness of the therapist during the parent-child interaction. PCIT, coupled with the DPICS, provides parents with quantitative feedback related to specific behaviors and broader behavioral patterns and interactions observed between the parent and child that may contribute to the development and maintenance of child behavioral problems; along with these data, the DPICS also provides a means to measure treatment progress and outcome (Brestan-Knight & Salamone, 2011; McMahon & Frick, 2005).

### **Psychometrics of the DPICS**

Many ABO systems are plagued by a lack of psychometric support, standardization, and ecological validity (Haynes, 2001; Mash & Foster, 2001; Mori & Armendariz, 2001; Roberts, 2001). Fortunately, most of these psychometric concerns have been addressed in the literature that supports the DPICS. The DPICS has been standardized with children ages 3 to 6 for normative and disruptive behavior disordered populations (Eyberg et al., 2005; Robinson & Eyberg, 1981). Studies also demonstrate that the DPICS has adequate inter-observer agreement, test-retest reliability, discriminative validity, convergent validity, and treatment sensitivity (see Eyberg et al., 2005 for a review; Bessmer, 1998; Bessmer & Eyberg, 1993; Brinkmeyer, 2006; Chaffin et al., 2004; Coursen, 2009; Deskins, 2005; Foote, 2000; McMahon & Frick, 2005; Robinson & Eyberg, 1981; Schuhmann et al., 1998; Webster-Stratton, 1985). However, much of the existing psychometric support for the DPICS, now in its third edition, has been extrapolated from studies using previous editions. Although it may be assumed that psychometric support generalizes across assessment editions, more studies are needed to bolster support for the updated coding system. For example, normative data are only available for children ages 3 through 7. Although normative data for children ages 8 through 12 were collected for

comparison children (Coursen, 2009) and for physically abused children (Deskins, 2005), more data are needed to extend the clinical utility of the DPICS-III to older populations (Eyberg et al., 2005).

A modified DPICS has also been developed by Webster-Stratton (1985) and used in clinic- and home-based observations of varying structure. Although Webster-Stratton conducted unstructured observations in the home, she did not do so using structured observations; thus, comparisons could not be made between structured clinic-based observations and structured home-based observations. Also, this study only analyzed composite behavioral categories. Therefore, any variability in individual codes may have been washed out when codes were combined, possibly erasing any differences that might have appeared between settings. Also, Pearson correlations were conducted instead of intraclass correlations, which might have limited statistical power (i.e., underestimated between-setting stability) in that the variability in codes attributable to coder differences may not have been taken into account when comparing codes across or within settings. Interestingly, comparisons between unstructured clinic-based and unstructured home-based observations found high correlations in composite categories; Webster-Stratton interpreted this as evidence that observation structure was more important than observation setting.

### **Dissemination of PCIT**

Recent dissemination efforts of PCIT into community clinics and home settings (Galanter, Self-Brown, Valente, Dorsey, Whitaker, Bertuglia-Haley, et al., 2012; Timmer, Zebell, Culver, & Urquiza, 2009; Ware, McNeil, Masse, & Stevens, 2008; Wilsie, Travis, Thornberry, Jr., & Brestan-Knight, 2010) have also coincided with attempts to improve the efficiency of the DPICS. For example, the DPICS codes have been refined with each new

edition, and an abridged manual has also been developed to maximize user-friendliness (Chase & Eyberg, 2006). Furthermore, recent studies suggest that the DPICS duration may be shortened without altering the behavioral data gathered by shortening or eliminating WU segments (Shanley & Niec, 2011; Thornberry & Brestan-Knight, 2011). Although these dissemination and utility-enhancing efforts are encouraging in the context of facilitating the use of EBA in a managed-care world, as has been discussed, there is a need to bolster psychometric support for ABOs. Specifically, in order to optimize confidence in the reliability and validity of the DPICS as applied in home-based PCIT, home-based DPICS standardization and reactivity studies are needed.

### **Study Goals**

The current study sought to provide additional psychometric support for the DPICS, in general, and assist current dissemination and implementation efforts by examining home-setting-specific psychometrics and the potential validity threat of reactivity. Specifically, this study:

1. Collected pilot standardization data for home-based DPICS observations with a non-clinical, community-based sample
2. Compared home-based DPICS standardization data to existing clinic-based DPICS pilot standardization data
3. Explored reactivity effects on child and parent behavior during video-recorded, clinic- and home-based DPICS observations using the TORQ (a newly developed measure of reactivity)

### **Hypotheses**

The specific hypotheses of the current study were:

1. Based on results from Webster-Stratton (1985), it was hypothesized that the only significant differences between home-based and clinic-based DPICS data would be parent praise, parent commands, and child noncompliance; specifically:
  - a. There would be significantly less parent labeled praise in home-based DPICS observations during all coding segments (CLP, PLP, and CU)
  - b. There would be significantly less parent unlabeled praise in home-based DPICS observations during all coding segments
  - c. There would be significantly fewer parent commands (direct commands, indirect commands, and no opportunity commands) in home-based DPICS observations during all coding segments
  - d. There would be significantly more child noncompliance in home-based DPICS observations during all coding segments
2. Based on results from Rhule, McMahon, and Vando (2009), it was hypothesized that on the TORQ:
  - a. Parents would report significantly higher representativeness of their and their child's behavior for and greater comfort with the CLP DPICS segment, followed by the CU segment, then PLP
  - b. Parents would report significantly more difficulty during the PLP segment, followed by CU, and CLP
3. Based on Gardner (2000), it was predicted that reactivity scores on the TORQ would be predicted by the number of observers present in a given observation such that more observers would be associated with more reactivity. However, it was hypothesized that other demographic variables would not predict differences in reactivity as measured by the TORQ.

4. Based on Kazdin (1982), it was hypothesized that higher reactivity scores on the TORQ would not predict child noncompliance or parent inappropriate behavior. However, it was predicted that higher reactivity scores would predict higher levels of prosocial behaviors, i.e., parent praise and child prosocial behavior as measured by the DPICS.



## Method

### Participants

Prior to advertising for this study, we obtained the approval of the university institutional review board. Subsequently, we posted flyers advertising this study in local groceries, businesses, pediatricians' offices, dentists' offices, daycares, and churches. Advertisements instructed interested families to contact the Auburn University Parent-Child Research Laboratory either by phone or email to acquire more information about the study. Only families of children between the ages of 2 and 10 were included in this study sample. Families were screened by phone and were not included in this study if they had previous contact with Child Protective Services. No families were excluded based on these criteria. Families were reimbursed \$20.00 USD and a small child's toy for participating in this study. A total of 32 families were recruited using the above procedure. Two families were excluded from subsequent analyses because their primary spoken language at home and during the DPICS observation was not English, preventing coders from reliably coding their interactions. Three families were excluded from data analyses because the children were scored by their parents in the clinically-significant range on at least one of the Behavior Assessment Scale for Children, Second Edition (BASC-2) composite scales (i.e., Externalizing, Internalizing, or Behavioral Symptoms Index) or on the Eyberg Child Behavior Inventory (ECBI). Thus, the current study included data from 27 families.

**Sample Demographics.** Demographic data were obtained from participating families using a demographics questionnaire completed by parents following completion of the DPICS

observation. Demographic data are presented by observation setting in Table 3. Overall, data were collected from 25 biological mothers and 2 biological fathers (mean age = 33.37,  $SD = 6.67$ ). These parents completed observations with 13 girls and 14 boys (mean age = 4.33,  $SD = 1.73$ ). Most of these families reported Caucasian ethnicities ( $n = 18, 66.7\%$ ), followed by African American ( $n = 7, 25.9\%$ ), Hispanic ( $n = 1, 3.7\%$ ), and Other ( $n = 1, 3.7\%$ ). The majority of the parents were married ( $n = 22, 77.8\%$ ), followed by single parents ( $n = 3, 11.1\%$ ), divorced parents ( $n = 2, 7.4\%$ ), and remarried parents ( $n = 1, 3.7\%$ ). Our sample was highly educated, with the majority reporting a Master's ( $n = 12, 44.4\%$ ) or Bachelor's ( $n = 6, 22.2\%$ ) degree in their fields. Six parents ( $22.2\%$ ) reported having some college experience, 2 ( $7.4\%$ ) reported having a doctoral degree, and 1 ( $3.7\%$ ) reported having an Associate's degree. Similarly, parents reported a high level of education for their spouses, with 7 ( $25.9\%$ ) reporting a Master's degree, 7 ( $25.9\%$ ) reporting a Bachelor's degree, 6 ( $22.2\%$ ) reporting a doctoral degree, 3 ( $11.1\%$ ) reporting some college experience, 1 ( $3.7\%$ ) reporting a high school education, and 3 ( $11.1\%$ ) caregivers not reporting their spouses' level of education. Families came from a variety of income ranges: \$0-10,000 = 2 families ( $7.4\%$ ), \$10,000-20,000 = 1 family ( $3.7\%$ ), \$20,000-30,000 = 5 families ( $18.5\%$ ), \$50,000-60,000 = 2 families ( $7.4\%$ ), \$60,000-70,000 = 4 families ( $14.8\%$ ), \$70,000-80,000 = 3 families ( $11.1\%$ ), \$80,000-90,000 = 3 families ( $11.1\%$ ), \$90,000-100,000 = 2 families ( $7.4\%$ ), and 5 families ( $18.5\%$ ) reporting greater than \$100,000 per year. Parents also reported a variety of occupations, from graduate students, teachers, and stay-at-home parents to administrators, counselors, and professors. Most families had a single child in the home ( $n = 13, 48.1\%$ ), followed by 2 children ( $n = 7, 25.9\%$ ), no children (i.e., child lives with other parent,  $n = 4, 14.8\%$ ), and 3 children ( $n = 3, 11.1\%$ ). Independent samples  $t$  tests were conducted to determine if there were any differences on these demographic variables between

families observed in the clinic and those observed in the home. No significant differences were found between groups with the exception of marital status; there were a significantly higher number of single parents in the home-based observation sample compared to the clinic-based observation sample,  $t(25) = 2.18, p = .04$ .

## **Measures**

**Demographics Questionnaire.** Participating caregivers completed a demographics questionnaire (see Appendix) that requested the following information: primary caregiver age, gender, race, marital status, level of education, occupation, field of work, and relationship with target child; spouse (if applicable) level of education, occupation, field of work, and relationship with target child; approximate yearly income for the household; target child age, gender, and race; and total number of children living in the home.

**Behavior Assessment Scale for Children – 2<sup>nd</sup> Edition Parent Rating Scale (BASC-2 PRS).** Parents were asked to complete the appropriate BASC-2 PRS (Reynolds & Kamphaus, 2004) based on the target child's age for screening purposes. The Preschool BASC-2 PRS is designed for caregivers with children aged 2- to 5-years-old, and the Child BASC-2 PRS is used with caregivers of children aged 6- to 11-years-old. The BASC-2 PRS measures child adaptive behavior and behavioral and emotional problems that occur in the home and community. The BASC-2 manual (Reynolds & Kamphaus, 2004) reports strong psychometric support, including strong internal consistency for the composite scales, a consistent factor structure, and convergent validity with other child behavior rating scales. The BASC-2 PRS has been standardized across age and gender, allowing for comparison of parent-reported behavioral/emotional problems with those of clinical and nonclinical populations. Caregivers answer items on the BASC-2 PRS by selecting one of four frequency scores (i.e., 0 = "Never," 1 = "Sometimes," 2 = "Often," and 3 =

“Almost Always”) that best reflects their perceptions of their child’s behaviors. Item raw scores are then converted into *T* scores such that higher scores on clinical scales indicate higher levels of problematic behaviors. For this study, BASC-2 PRS reports were used to screen participants for clinically-significant levels of problematic behavior. Only families that reported subclinical levels of behavioral problems (i.e., *T* scores < 70) on the Externalizing Problems, Internalizing Problems, and Behavioral Symptoms Index composites were included in this study. The validity scales of the BASC-2 PRS were also used as a measure of socially-desirable responding.

**Eyberg Child Behavior Inventory (ECBI).** The ECBI (Eyberg & Pincus, 1999) is a 36-item, narrow-band parent report measure of child disruptive behavior. Items on the ECBI relate particularly to behaviors characteristic of Oppositional Defiant Disorder and Attention-Deficit/Hyperactivity Disorder diagnoses. The ECBI consists of two scales: the Intensity scale and Problem scale. The Intensity scale is calculated by adding parent responses on 7-point Likert type items where higher scores indicate higher frequencies of behavioral problems. Problem scale scores are calculated by totaling the number of Yes-No items on which parents identify the child’s behavior as being problematic for them. Raw scores on the two ECBI scales can be converted into T-scores; however, typical clinical application of the ECBI during PCIT uses scale raw scores as clinical cutoffs (i.e., Intensity: 131; Problem: 15) and a criterion for treatment completion (Intensity  $\leq$  114 or  $\frac{1}{2}$  a standard deviation above the normative mean; Eyberg, 2010).

The ECBI was restandardized in 1999 with an outpatient sample of children aged 2 to 16. Its manual documents adequate internal consistency for both the Intensity and Problem scales (coefficients of .95 and .93, respectively) as well as 3-week test-retest reliability (.86 and .88, respectively). There is also evidence of convergent validity in that the ECBI correlates with other instruments such as the Child Behavior Checklist (Achenbach & Rescorla, 2001) and the

Parenting Stress Index (Abidin, 1995). The ECBI has also been shown to discriminate between clinic-referred and non-referred children and to be sensitive to treatment effects brought about by parent training programs such as PCIT (Eyberg & Pincus, 1999). The ECBI was used to exclude families with clinically-significant behavioral problems from the study (defined as having clinically-significant scores on both the Intensity and Problem scales).

**Thornberry Observation Reactivity Questionnaire (TORQ).** The TORQ was developed for this study to investigate reactivity effects on parent and child behavior during each segment of the DPICS observation. This questionnaire includes 16 Likert-type items and 13 free-response items. Likert-type items have a 4-point anchor format that allows for extreme answers (e.g., “Very Difficult”) and moderate answers (e.g., “Somewhat Similar”). The Likert-type items produce four scales: Parent Behavior Representativeness, Child Behavior Representativeness, Parent Perceived Difficulty, and Parent Comfort. Each scale can be assessed for each coding segment of the DPICS observation (i.e., CLP, PLP, CU). These scales can also be summed across the DPICS coding segments to produce a total score for the entire DPICS observation such that lower total scores on each scale indicate greater representativeness of child and parent behavior, greater ease in conducting the observation tasks, and greater comfort with the observation (after the Parent and Child Representativeness and Difficulty scores are reverse-scored). This scoring system produces a minimum and maximum score (that is, the least and most reactive score, respectively) of 4 and 16 for each DPICS segment and a minimum and maximum overall reactivity score of 16 and 64, respectively. Likert-type items related to parent and child behavioral representativeness and perceived difficulty of the observation task are followed by free-response items in order to determine what behaviors or what aspects of the

observation task may have contributed to unrepresentative behavior or hurt the external validity of DPICS observations. See Appendix for items and format of the TORQ.

**Dyadic Parent-Child Interaction Coding System (DPICS-III).** The abridged version of the DPICS-III (Chase & Eyberg, 2006) was used to code video-recorded DPICS behavioral observations collected in participants' homes or the Auburn University Psychological Services Center. The abridged DPICS-III collects frequency counts of various child and parent behaviors (e.g., Prosocial Talk (PRO), Command (CM), Labeled Praise (LP), Unlabeled Praise (UP), Neutral Talk (TA)). These categories of behaviors can then be combined using formulae set forth by Eyberg et al. (2005) to create composite categories. For this study, the composite categories for child behavior included Compliance, Noncompliance, and Inappropriate Behavior, and the composite categories for parent behavior included Inappropriate Behavior and Prosocial Behavior. For a list of child and parent behavior codes, see Table 1. For a list of child and parent composite categories and their formulae, see Table 2.

### **Procedure**

Upon initial contact by interested families, a laboratory member provided a brief description of the study procedures and reimbursements provided for participants' time and effort during the study. If still interested in participating in the study, families were scheduled for either an in-home or in-clinic data collection visit. Trained observers were assigned to either meet the family in the clinic or visit families' homes, where they obtained written parent consent, conducted a standardized DPICS-III observation, and administered study measures. Observers also video-recorded the DPICS observation, which was later coded by trained DPICS coders in the laboratory. Immediately following the DPICS observation, caregivers were asked to complete the TORQ (to ensure the observation was fresh in the caregiver's mind), followed by

the rest of the study measures. Following completion of measures, families were compensated as described above.

**Observers.** Undergraduate research assistants were recruited and trained as observers. Observers were trained to administer study measures and conduct and videotape the DPICS observations using standardized instructions and practice scripts. Observers were encouraged to minimize their interactions with family members while conducting the DPICS observations. However, if observers were engaged directly by caregivers or children during the DPICS observation, they were instructed to respond briefly. This writer or another graduate student accompanied observers during their first observation to ensure observations were reliably collected. In situations during which observers were unable to collect data due to scheduling conflicts, this writer or another graduate student collected data.

**Training of Coders.** Undergraduate research assistants who were designated as coders for this project completed a rigorous training process that included the completion of the DPICS-III workbook (see Eyberg et al., 2005), attendance of regular practice meetings led by a faculty supervisor or graduate student trained in the DPICS, and reliable coding of several criterion video-recordings. Weekly practice meetings consisted of checking coders' progress with the workbook, answering questions related to coding, and coding practice video-recordings or role-playing situations under the supervision of the faculty supervisor or graduate students. Upon completion of the workbook, coders were required to code criterion observations with a reliability of at least 80% agreement on 2 separate observations.

**Coding Procedures.** Video-recorded observations were randomly assigned to be coded by a team of DPICS-III trained coders who were blind to the study's hypotheses. Only coders who successfully completed the training procedures listed above were permitted to code

observations for this study. As coders viewed the recorded observations, they made tally marks for each occurrence of specific parent and child behaviors, as defined by the abridged version of the DPICS-III manual (Chase & Eyberg, 2006), on one of two coding sheets (see Appendix). All recorded segments were watched twice by coders, once to observe the child's behaviors and once to observe the parent's behaviors. One-third of the collected observations were coded by this writer as a reliability check. Frequency count totals obtained by the undergraduate research assistants were entered into a computer database and compiled into the various composite categories for statistical analyses.



## Results

**BASC-2 Descriptives.** Caregivers completed the age-appropriate BASC-2 form for their child. This resulted in 22 preschool and 5 child-aged forms. All responders had valid measures, as indicated by “Acceptable” scores on the F, Response Pattern, and Consistency validity scales. Only one exception to this occurred, with one family scoring in the “Caution” range on the Consistency scale, but this family had acceptable scores on the other two validity scales; thus, their responses were included in these data. Average BASC-2 scores across the various clinical scales and composites are presented by observation setting in Table 4. Average scores for all BASC-2 composites and subscales for both observation groups fell in the normal range. Based on the results of planned orthogonal pairwise contrasts, both groups had comparable scores on the various BASC-2 composites and subscales, with the exception of the Aggression scale; this scale was statistically significantly higher in the clinic-based group as compared to the home-based observation group,  $F(1, 25) = 6.36, p = .02$ .

**ECBI Descriptives.** All parents were able to complete the ECBI, the results of which are summarized in Table 4. One family reported ECBI Intensity scores in the clinical range (i.e., >131) but their Problem scale was in the normal range (i.e., <15); therefore, their data were included in these analyses; all other families reported ECBI scores in the normative range. ECBI Intensity scores ranged from 52 to 141, and the average ECBI Intensity score for the entire sample was 95.00 ( $SD = 25.06$ ). The range of the Problem scale was from 0 to 24, and the average Problem score was 7.52 ( $SD = 6.23$ ). As shown in Table 4, there were no significant

differences in ECBI Intensity or Problem scales between clinic-based and home-based observation samples.

**TORQ Descriptives.** TORQ results are summarized in Table 5. Internal consistency for the TORQ was good, with a Cronbach's alpha of .78. Kolmogorov-Smirnov tests of normality found that many TORQ scales were normally distributed (i.e.,  $p > 0.05$ ; Total Reactivity, Parent Representativeness, PLP Total, and CU Total Reactivity). However, it was also found that the Child Representativeness and Comfort scales were positively skewed, and the CLP Total Representativeness scale was found to be platykurtic and the Child Representativeness scale was found to be leptokurtic.

Table 6 illustrates bivariate Pearson correlations between TORQ subscales. Looking at correlations between total reactivity scores by DPICS segment (i.e., CLP Total, PLP Total, and CU Total), it is apparent that there are significant, positive relationships between these scales, indicating that high reactivity during one segment of the DPICS likely coincides with high reactivity throughout the DPICS observation. Not surprisingly, all scales have a strong, positive correlation with the Total Reactivity scale of the TORQ. Importantly, not all subscales of the TORQ correlated significantly, indicating that some subscales, such as the Parent Reactivity, Child Reactivity, Difficulty, and Comfort scales, are measuring different components of reactivity.

A series of paired samples  $t$  tests were conducted to compare total reactivity scores across the various DPICS segments (i.e., CLP Total, PLP Total, and CU Total). Effect sizes were also calculated. These analyses found that total TORQ-measured reactivity during CLP was significantly lower than total reactivity during PLP,  $t(26) = -3.81, p = .001, d = -0.83$ , and significantly lower than total reactivity during CU,  $t(26) = -2.35, p = .03, d = -0.51$ . However,

total reactivity scores during PLP and CU were not significantly different,  $t(26) = 1.37, p = .18, d = 0.25$ . Another paired samples  $t$  test was conducted to examine differences in behavioral representativeness between parents and children across all DPICS segments, as measured by the TORQ (i.e., Parent Representativeness and Child Representativeness scales); this comparison found that parents reported more total reactivity throughout the entire observation affecting their own behavior (i.e., higher scores on the Parent Representativeness scale) than their child's behavior (i.e., Child Representativeness scale),  $t(26) = 3.86, p = .001, d = 0.86$ .

To answer Hypothesis 2, chi-squared analyses were used to determine if reactivity scores on the various TORQ scales (i.e., Child Representativeness, Parent Representativeness, Comfort, Difficulty) varied by DPICS observation segments (i.e., CLP, PLP, and CU). These analyses found no differences in the proportions of responses on the Child Representativeness and Parent Representativeness scales across DPICS segment. However, there were significant differences in the proportions of responses across DPICS segments on the TORQ Comfort scale,  $\chi^2(6, 81) = 16.14, p = .01$ , and on the TORQ Difficulty scale,  $\chi^2(6, 81) = 18.50, p = .005$ . Specifically, 19 families reported feeling Very Comfortable during CLP, 17 reported feeling Very Comfortable during CU, and only 8 reported feeling Very Comfortable during PLP. Similarly, 23 families reported that following the rules of CLP was Very Easy, 11 reported the rules of CU was Very Easy, and 9 families reported the rules of PLP were Very Easy.

Orthogonal planned pairwise contrasts were also performed on the various TORQ scales in order to compare reactivity between observation settings. These results and effect sizes are summarized in Table 7. Overall, there were no statistically significant differences on any TORQ scale between clinic and home observation settings.

**Regression of Demographic Variables on TORQ Scales.** To answer Hypothesis 3, a series of multiple linear regressions were conducted to determine if sample demographic variables predicted variability in TORQ scale scores. Using a stepwise entry method, the following demographic variables were tested to see if they predicted a significant amount of variance in TORQ scales: child age and gender; ethnicity (parent and child ethnicity were identical for all observations; thus, these variables were collapsed); parent age, gender education, and marital status; number of children in the home; and yearly family income. Additionally, the number of observers present during the observation (1 or 2) was added to the regression model to determine if this predicted a significant amount of unique variance in TORQ scale scores. For TORQ Total scores, ethnicity was a significant predictor,  $F = 6.01, p = .02, Adj. R^2 = .16$ , with African American families reporting lower reactivity than Caucasian families. None of the above variables entered into the regression model significantly predicted variance in CLP Total Reactivity on the TORQ. For PLP Total Reactivity, ethnicity again predicted a significant amount of variance,  $F = 11.93, p = .002, Adj. R^2 = .30$ , such that African American families reported less reactivity during PLP than Caucasian families. Variance in CU Total Reactivity scores on the TORQ was not predicted by any variable entered into the regression. Similarly, Child Representativeness and Comfort scores on the TORQ were not predicted by demographic variables or number of observers. Parent Representativeness scores were predicted by parent age,  $F = 9.37, p = .005, Adj. R^2 = .24$ , with older parents reporting more reactivity in their own behavior. Finally, TORQ Difficulty scores were predicted by parent and child age. Specifically, parent age predicted approximately 11% of adjusted variance in Difficulty scores ( $F = 4.30, p = 0.05, Adj. R^2 = .11$ ), with older parents reporting more difficulty during the entire DPICS observation. Adding child age to the model predicted significantly more variance in the

Difficulty TORQ scale ( $F = 4.52, p = .04, Adj. R^2 = .22$ ). Controlling for the effects of parent age, perceived difficulty was reportedly higher with younger children. Number of observers did not predict any TORQ scale.

**DPICS Reliability Analyses.** DPICS inter-rater reliability was assessed on a random sampling of 33% of video-recorded observation segments by the primary investigator. Thirty segments were coded by the primary investigator, 15 from in-home and 15 from in-clinic observations. Of the 15 segments in each observation setting group, 5 each were randomly selected from CLP, PLP, and CU segments. Thus, inter-rater reliabilities equally represent all DPICS segments across both observation settings. However, due to 3 families being removed from this study because of clinically-significant scores on screening measures, DPICS reliability results only reflect the remaining 27 families. Inter-rater reliability was calculated using intraclass correlations (ICCs) with one-way, random effects models and average measurement reliability by comparing frequency counts of a given DPICS code obtained by DPICS coders with those obtained by the primary investigator (see Table 8).

ICC values were calculated for separate observation settings (i.e., home- versus clinic-based) and for the entire combined sample of 27 families. For parent DPICS codes, ICCs for the entire sample ranged from .68 to .98. For the clinic-based observations, parent ICCs ranged from .64 to .97. For home-based observations, values ranged from .84 to .99. For child codes, ICCs for the entire sample ranged between .67 and .99. For clinic-based observations, values ranged from .67 to .99, and, for home-based observations, ICCs varied between .68 and .98. According to criteria by Shrout and Fleiss (1979), ICC coefficients are described as acceptable when they equal or exceed .75. Based on these criteria, 20 of the 22 codes used in this study were considered satisfactory for the full sample. Child No Answer and Parent Behavior Description

were the only codes below this cutoff for the entire sample. For clinic-based observations, 16 of the 22 codes exceeded Shrout and Fleiss' (1979) .75 cutoff. Codes below this value included Child Noncompliance, Child No Answer, Child No Opportunity to Answer, Parent Labeled Praise, Parent Reflection, and Parent Behavior Description. Finally, for home-based observations, 20 of the 22 codes were coded with acceptable interrater reliability (i.e., ICC > .75). Only Child No Answer fell below this value. Notably, reliability for the Behavior Descriptions code was incalculable for home-based observations as this code was never observed in this setting.

Fisher r-to-Z transformations were conducted to compare ICCs for the various DPICS codes between observation settings. Results of these analyses are summarized in Table 8. Significant differences emerged for some DPICS codes. Specifically, Child Whine was coded significantly more reliably in the clinic-based video-recordings as opposed to the video-recordings from the home setting,  $Z(25) = 2.15, p = .03$ . In contrast, Child Noncompliance was significantly more-reliably coded in the video-recordings from the home setting,  $Z(25) = 3.29, p < .001$ , as was Child No Opportunity to Comply,  $Z(25) = 2.35, p = .02$ . There were no statistically-significant differences in the coding reliability of parent codes between observation settings.

**In-Home Versus Clinic-Based DPICS Comparisons.** To answer Hypothesis 1, DPICS-coded parent and child behavioral frequency count means were compared across observation settings (clinic and home) using planned orthogonal pairwise contrasts. Results are separated by observation segment (i.e., CLP, PLP, and CU) and can be found in Table 9, Table 10, and Table 11. During CLP, no DPICS code significantly differed in mean frequency across observation setting. During PLP segments, parents did not differ in their DPICS-coded behavior between

settings. However, children differed in Command frequencies, with children giving statistically significantly more commands in the clinic as compared to the home setting,  $F(1, 25) = 5.49, p = .03, d = 0.95$ . The most significant differences between parent and child behaviors by observation setting occurred during CU. Here, children again used more commands in the clinic than in the home-based observation,  $F(1, 25) = 7.08, p = .01, d = 1.07$ . Also, children asked more questions in the clinic,  $F(1, 25) = 9.57, p = .005, d = 1.24$ . They also provided more answers to parents' questions in the clinic when compared to the home,  $F(1, 25) = 6.55, p = .017, d = 1.03$ . However, parents did not ask significantly more questions (IQ or DQ) in the clinic. Again, parents did not differ on any DPICS codes between settings.

**Regression of TORQ Scales on the DPICS Composite Categories.** Finally, to answer Hypothesis 4, a series of regressions were calculated to determine if TORQ scales significantly predicted variance in DPICS composite categories. To boost statistical power, and because TORQ scores did not significantly differ between observation settings, data from in-home and in-clinic observations were combined for these regression analyses. Using a stepwise entry method for these regressions, models were created for the five DPICS composite categories listed in Table 2. These composites were calculated for each DPICS observation segment (i.e., CLP, PLP, and CU) and total composites were created for the entire DPICS observation by summing (Parent Inappropriate, Parent Prosocial, Child Prosocial, and Child Inappropriate) or averaging (Child Compliance and Child Noncompliance Ratios). Demographic variables were also entered into the regression models, followed by TORQ scales to determine if TORQ scales predicted variance in DPICS composite categories above that predicted by demographics.

Results of these regressions showed that, during the CLP segment, Parent Prosocial behaviors (i.e., a sum of Unlabeled Praise, Labeled Praise, Behavior Descriptions, and

Reflections) were significantly predicted by two scales of the TORQ: CLP Total Reactivity and PLP Total Reactivity. CLP Total Reactivity predicted approximately 27% of the variance in these prosocial behaviors ( $F = 10.81, p = .003, Adj. R^2 = .27$ ), and adding PLP Total Reactivity to the model predicted significantly more variance ( $F = 8.25, p = .002, Adj. R^2 = .36$ ).

Specifically, higher perceived total reactivity during CLP and PLP was associated with more prosocial behaviors during CLP. No TORQ scales significantly predicted Parent Inappropriate or Child Inappropriate behaviors during CLP. However, Parent Inappropriate behaviors were significantly predicted by marital status ( $F = 5.14, p = .03, Adj. R^2 = .14$ ) such that married parents had a higher mean frequency of inappropriate behaviors, followed by divorced parents, remarried parents, and single parents. TORQ scales and demographic variables did not predict Child Compliance or Child Noncompliance Ratios during CLP. However, PLP Total Reactivity predicted a significant amount of variance in Child Prosocial behaviors during CLP ( $F = 7.07, p = .01, Adj. R^2 = .19$ ) such that more reactivity was associated with more prosocial child behaviors.

For the PLP segment of DPICS observations, Parent Prosocial behavior was predicted by Total Reactivity scores during CU,  $F = 12.10, p = .002, Adj. R^2 = .30$ , such that higher total reactivity during CU was associated with more prosocial behaviors by parents during PLP. Adding the number of observers to the model predicted significantly more variance in Parent Prosocial behaviors ( $F = 12.65, p < .001, Adj. R^2 = .47$ ) such that more observers was associated with fewer prosocial behaviors during PLP. In addition, adding ethnicity to the model predicted significantly more variance ( $F = 12.00, p < .001, Adj. R^2 = .56$ ). Specifically, African American parents on average had fewer prosocial behaviors during PLP than did other ethnicities. TORQ-reported Comfort predicted Child Prosocial behavior during PLP,  $F = 5.72, p = .025, Adj. R^2 =$



.154, with higher levels of parent discomfort being associated with higher frequencies of child prosocial behavior. TORQ scales and demographic variables did not successfully predict Parent Inappropriate or Child Inappropriate behaviors. Ethnicity significantly predicted Child Compliance during PLP ( $F = 4.39, p = .05, Adj. R^2 = .12$ ) such that African American children on average were less compliant than other ethnicities during PLP. This model predicted significantly more variance in Child Compliance when number of children living in the home was added,  $F = 5.24, p = .01, Adj. R^2 = .25$ , with more children being associated with higher compliance. Adding the Parent Representativeness scale of the TORQ predicted significantly more variance ( $F = 6.01, p = .004, Adj. R^2 = .37$ ), such that, when ethnicity and number of children in the home were controlled for, lower parent-reported reactivity of their own behavior was associated with higher child compliance. Interestingly, TORQ scales and demographic variables did not predict Child Noncompliance Ratios during PLP.

During CU, Parent Prosocial behaviors were significantly predicted by numerous variables. Using the stepwise regression procedure, the TORQ CLP Total Reactivity scale ( $F = 9.43, p = .005, Adj. R^2 = .245$ ) was the first variable entered, with higher reactivity during CLP being associated with higher frequencies of parent prosocial behaviors during CU. Adding marital status predicted significantly more variance in CU Parent Prosocial behaviors ( $F = 8.30, p = .002, Adj. R^2 = .36$ ), with married parents displaying the highest mean frequency of prosocial behaviors, followed by divorced parents, remarried parents, and single parents. This model predicted significantly more variance when number of observers was entered ( $F = 9.55, p < .001, Adj. R^2 = .50$ ) such that, when the effects of CLP Total Reactivity and marital status were controlled, fewer observers was associated with higher frequencies of prosocial parenting behaviors during CU. The regression model predicted significantly more variance in CU Parent

Prosocial behaviors when parent ( $F = 10.62, p < .001, Adj. R^2 = .60$ ) and child gender ( $F = 12.91, p < .001, Adj. R^2 = .70$ ) were added to the model. Specifically, fathers tended to use fewer prosocial behaviors during CU and parents used fewer prosocial behaviors with boys. Parent Inappropriate behaviors during CU were not predicted by TORQ scales, but they were predicted by number of children living in the home ( $F = 5.72, p = .025, Adj. R^2 = .15$ ) such that more children was associated with more inappropriate behaviors. During CU, Child Prosocial behavior was predicted by the Comfort scale of the TORQ ( $F = 12.67, p = .002, Adj. R^2 = .31$ ) such that higher discomfort was associated with higher child prosocial talk during CU. This model predicted significantly more variance in child prosocial talk when family income was entered into the model ( $F = 11.61, p < .001, Adj. R^2 = .45$ ) with higher child prosocial talk being associated with higher family income when comfort was controlled. Significantly more variance in child prosocial talk was predicted by this model when the CU Total Reactivity scale of the TORQ was entered ( $F = 12.53, p < .001, Adj. R^2 = .57$ ). When the effects of TORQ-measured discomfort and family income were controlled, higher CU Total Reactivity scores were associated with lower average frequency of child prosocial talk. TORQ scales did not predict Child Inappropriate behaviors during CU but number of children living in the home did ( $F = 5.34, p = .03, Adj. R^2 = .14$ ), such that more children was associated with higher frequencies of inappropriate behaviors. Similarly, Child Compliance and Child Noncompliance were not predicted by TORQ scales but were predicted by child age (both  $F = 7.84, p = .01, Adj. R^2 = .22$ ) with older children tending to have higher compliance and lower noncompliance ratios.

For the last set of regression analyses, DPICS frequency counts were summed (or averaged in the case of the compliance ratios) across all DPICS coding segments. Using this method, Parent Prosocial behavior during the entire DPICS observation was significantly

predicted by TORQ Total Reactivity ( $F = 14.24, p = .001, Adj. R^2 = .337$ ), with higher levels of reactivity being associated with higher levels of parent prosocial behavior. A significantly higher portion of variance was predicted when number of observers was entered into the regression ( $F = 11.54, p < .001, Adj. R^2 = .45$ ), with more observers being associated with fewer average parent prosocial behaviors. Adding child gender to the model significantly increased variance predicted yet again ( $F = 11.19, p < .001, Adj. R^2 = .54$ ) with boys being associated with fewer parent prosocial behaviors throughout the entire DPICS observation. Marital status added significantly to the regression model ( $F = 11.32, p < .001, Adj. R^2 = .61$ ), with married parents on average displaying more prosocial behaviors, followed by divorced, remarried, then single parents. Finally, Parent Prosocial behavior was best predicted when the CLP Total Reactivity scale of the TORQ was added to the model ( $F = 13.30, p < .001, Adj. R^2 = .70$ ). When the other variables in the model are controlled for, higher total reactivity during CLP was associated with higher average frequencies of positive parent behaviors during the entire DPICS observation. Both total Parent Inappropriate behaviors and Child Inappropriate behaviors during the entire DPICS were not predicted by any TORQ scale or demographic variable. Child Prosocial Talk during the entire DPICS observation was significantly predicted by the TORQ Comfort scale ( $F = 9.85, p = .004, Adj. R^2 = .25$ ) with higher parent-reported discomfort being associated with higher frequencies of child prosocial talk. Finally, Child Compliance and Noncompliance Ratios averaged across the entire DPICS observation were not predicted by any TORQ scale or demographic variable in this study.

## **Discussion**

The EBPP movement has gained momentum in recent years. This progress is illustrated by an increase in efforts to establish EBA guidelines, an initiative which has been relatively neglected compared to EBT research (Achenbach, 2005; Hunsley & Mash, 2010, 2011; Jensen-Doss, 2011). This new wave of interest in EBA is perhaps best exemplified by work related to the EBA of children and families and a subsequent recommitment to including behavior observations in this discussion (Achenbach, 2005; Cashel, 2002; Mash & Hunsley, 2005). As a consequence of this new focus on behavior observations, researchers have become aware of the complacency that exists amongst clinicians who occasionally ignore the psychometrics, good or bad, of observations (Aspland & Gardner, 2003; Gardner, 2000; Haynes, 2001; Mash & Foster, 2001; Roberts, 2001). In order for EBPP to continue positive progress, the weaknesses of our assessment practices cannot be ignored any longer.

This study sought to accomplish several goals in the interest of strengthening the psychometric support of the DPICS, a standardized ABO used to measure parent and child interactions. First, this study was the first of its kind to collect home-based observations with community families and compare these observations to observations conducted in a clinic setting with a similar population. Second, this study is the first of its kind to reexamine the concern of reactivity during behavior observations using an as-yet unused method: parent report of self and child reactivity using the TORQ. This study is also the first to consider parent-reported comfort and perceived difficulty in a conceptualization of reactivity during observations. Finally, this

study is the first to attempt to predict variance in child and parent behavior during the DPICS using a measure of reactivity.

Collecting normative DPICS data is an important first step in disseminating PCIT to home-based treatment settings. As Haynes (2001) points out, the “dynamic nature of validity” dictates that researchers and clinicians not assume that the psychometric qualities of a measure transfer to new settings. As this study found, there were important differences between DPICS data collected in the home and the clinic. Indeed, interrater reliability varied significantly between observation settings (i.e., home and clinic) for some codes, including child whine, noncompliance, and no opportunity to comply. Only whine was coded with higher reliability in the clinic setting; the remaining codes were coded with higher reliability in the home setting. This finding runs counter to anecdotal evidence of coding difficulties in home-based observations, such as noisy environments (e.g., dishwashers, air conditioners) and poorer video quality. It is possible that coders attended better to home-based observations due to the novelty of these settings as compared to the standardized, less-stimulating clinic setting. It is also possible that codes were coded with varying reliability due to their low base rate. For example, this normative sample of children was rarely noncompliant, and parents rarely used some of the highly-specialized behaviors that are trained during in vivo PCIT sessions over a period of months. Importantly, these data illustrate that DPICS observations can be conducted in the home setting and be coded reliably, a vital finding if PCIT is to be disseminated into home-based treatment models.

This study also found behavioral differences in children but not parents between observation settings. These findings should be considered tentative given this small, normative sample and should not be assumed to generalize to clinical populations. Contrary to our

hypotheses, there were no differences in parent praise or child noncompliance during any DPICS segment. However, children were more commanding during PLP when observed in the clinic as compared to the home. It is possible that children were more uncomfortable in the novel, clinic environment during an unfamiliar play situation than were children engaging in the same play situation in their own home, leading to attempts to control the parent during clinic-based PLP. Notably, most differences in child behavior between observation settings occurred during the CU portion of the DPICS observation. At this time, children tended to use more commands, ask more questions, and answer parents' questions more frequently (although not significant, parents tended to ask more questions during clinic-based CU segments). Also during CU, parents tended to talk more in the clinic setting, averaging 10 more neutral verbalizations in this five-minute segment than parents in the home. However, this difference in parent behavior was not statistically significant. Given Roberts' (2001) statement that clean-up tasks appear to be the most useful and valid measures of parent-child behavior, the differences we found during this segment of the DPICS between observation settings appears all the more important to consider in the contexts of clinical and empirical work. Although these behavioral differences between observation settings were rare in this normative sample, these findings support Haynes' (2001) recommendation that psychometric characteristics of assessments not be assumed to transfer to different settings, populations, etc.

Interestingly, the few observed differences in parent and child behavior across observation settings cannot be entirely attributed to reactivity, as TORQ reactivity scores in this normative sample did not significantly differ across observation setting. Furthermore, differences in parent and child behavior across coding segments (CLP, PLP, and CU) cannot be fully explained by varying levels of reactivity as total reactivity did not significantly differ between

DPICS segments where behavior did. Thus, independent of the effects of reactivity, parent and child behavior appear to differ across observation setting as a function of the varying task demands present during the three analogs. These results are not surprising given past research that shows behavioral differences across analog tasks (Webster-Stratton, 1985). However, the specific behaviors found to vary between observation settings in this project differ from those found to differ in previous studies. Namely, this study found no significant differences in child noncompliance, parent praise, or parent commands across observation settings. These findings may be sample-specific in that variance in child and parent behavioral frequencies may have been restricted due to this being a normative sample. Future studies are needed to compare parent and child behaviors across observation settings using normative and clinical populations.

The present study found mixed support that TORQ-measured reactivity varied across DPICS segments. Total Reactivity was lowest during CLP but did not differ significantly between PLP and CU. Contrary to our hypotheses, the parent-reported representativeness of child and parent behavior did not significantly differ across DPICS segments. However, parents reported significantly less representativeness of their own behavior for the entire DPICS as compared to the representativeness of their child's behavior. Taken together, these results suggest that parents observed differences in their own behavior during the entire DPICS more than they observed differences in their child's behavior when compared to how they and their child would typically act at home. However, there was not a single DPICS segment parents perceived as significantly more or less representative.

As predicted, parent reported comfort and difficulty varied with DPICS segment such that more parents reported higher levels of comfort and lower levels of difficulty during CLP, followed by CU, then PLP. According to previous research (Aspland & Gardner, 2003; Gardner,

2000; Kazdin, 1982), we might expect reactivity or discomfort to be highest during initial observation segments due to the novelty of the task, uncertainty of how the observation will proceed, etc., and would expect parents to become more comfortable over time as they habituate to the observation task. However, this study found that parent-reported comfort varied by DPICS segment such that more parents reported feeling very comfortable during CLP, followed by CU, then PLP. These findings may reflect Roberts' (2001) suggestion that PLP analogs are more unfamiliar, less comfortable, more difficult, and less informative than other ABO analogs. We might also expect reactivity to vary with the obtrusiveness of the assessment (Gardner, 2000), but this study found, contrary to our hypotheses, no direct effect of number of observers on reactivity scores. Furthermore, one might expect in-home observations to be more obtrusive than clinic-based observations, but this was not reflected by TORQ-measured differences in reactivity between these settings. It is possible that observations were not sufficiently obtrusive to increase TORQ-measured reactivity given a lack of variability in number and stimulus quality of observers (i.e., all observers were college or graduate students). Future studies with fewer (0, 1) or more (>3) observers or observers of varying stimulus value (e.g., psychologists, Child Protective Services workers) might vary obtrusiveness in a manner reflected in parent report measures of reactivity. Overall, these findings support the need to measure parents' comfort and perceived difficulty when considering the impact of reactivity on their and their children's behavior during an observation. Also, future studies with measures like the TORQ can assess the effects of obtrusiveness and habituation on reactivity.

Contrary to Hypothesis 3, TORQ scales were predicted by demographic variables. Specifically, variance in TORQ Total Reactivity across the entire DPICS was predicted by ethnicity such that African American parents reported on average less reactivity than did



Caucasian families. The same was true for Total Reactivity during PLP but not for CLP or CU. Taken together, these results suggest that the African American parents in this study perceived less reactivity throughout the DPICS observation, particularly when they were in charge of the play. These differences cannot be attributed to parent or child representativeness, comfort, or difficulty, as these scales were not predicted by ethnicity. Future studies are needed to investigate the effects parent and child ethnicity may have on reactivity. These results may ultimately prove to be unimportant given the questionable validity of PLP analogs (Roberts, 2001) and should be considered tentative given this small, normative sample.

In addition to ethnicity, parent age predicted variance in Parent Representativeness and Difficulty on the TORQ. Specifically, older parents reported more reactivity via less-representative behaviors in themselves and more difficulty during the DPICS observation. Child age also predicted Difficulty scores in that observations with younger children were reported as being more difficult by parents. These results suggest there may be cohort effects on reactivity such that older parents may be more perceptive of their behavior and may also have more difficulty when playing with their child, especially younger, more active children who require more redirection or limit-setting. These results have implications for conducting behavior observations with parents and children of varying ages and illustrate the importance of considering age when studying reactivity in the future.

Overall, the predictive power of the TORQ coincided with our hypotheses, which were based on the expectations set forth by Kazdin (1982) thirty years ago. That is, Kazdin's review suggests that reactivity may have a larger impact on positively-valenced (e.g., parent praise) rather than negatively-valenced (e.g., criticism) behaviors. In our study, TORQ scales did not predict inappropriate child or parents behaviors during any segment of the DPICS, but scales did

predict child and parent prosocial talk throughout the DPICS. Contrary to our original hypotheses, child compliance during PLP was predicted by the Parent Representativeness scale of the TORQ, but only after ethnicity and number of children living in the home were considered first. Given that PLP is often referred to as the “strangest” ABO task for families (Roberts, 2001) and produced the highest overall reactivity scores of the three DPICS segments in this study with our normative sample, it is possible that reactivity does play a part in child compliance during behavior observations and that future studies with populations exhibiting higher levels of inappropriate behavior may display differential predictive utility of the TORQ. However, more studies with larger, clinical populations are needed to fully understand the role of reactivity in predicting parent and child behavior during observations.

The results of this pilot study hold important implications for the field in terms of clinical work and additional research questions that require more studies. First, given the recent growth in dissemination of PCIT and home-based clinical services (Galanter, Self-Brown, Valente, Dorsey, Whitaker, Bertuglia-Haley, et al., 2012; Timmer, Zebell, Culver, & Urquiza, 2009; Ware, McNeil, Masse, & Stevens, 2008; Wilsie, Travis, Thornberry, Jr., & Brestan-Knight, 2010), this study highlights the importance of gathering norms during home-based observations as differences in parent and child behavior can occur between this setting and the clinic. Although a few differences were found in child behaviors between observation settings, the majority of DPICS coded behaviors did not differ. Thus, behaviors observed in the home may be more comparable to those observed in the clinic than they are different. Future studies should replicate these findings with larger and more representative samples and should seek to do the same with clinical samples, for whom these setting differences may be exaggerated or qualitatively different. Particular populations that should be studied in this context include

families with children diagnosed with disruptive behavioral disorders. Although PCIT has been shown to produce long-term, generalizable, positive change in this population (e.g., Funderburk et al., 1998; McNeil et al., 1991), gathering home-based observational data may further improve these treatment effects by allowing clinicians to more-effectively tailor treatment to target specific parent and child behaviors as they occur in the home. Child maltreatment populations may also uniquely benefit from studies similar to this project given that child victims of neglect, a difficult population to study, measure, and treat, have been found to benefit from home-based services (Gillaspy & Bonner, 2010).

Our emphasis on the need to collect normative behavioral observation data in the home setting should not be misconstrued as an indictment against clinic-based observations. Clinicians have been using clinic-based observational data effectively since the origins of clinical child psychological treatment. However, the differences in parent and child behaviors between observation settings in this study beg the question of whether clinic-based observations gather the full story of what is occurring in the parent-child relationship at home. It is possible that certain stimuli in the home environment (e.g., particular toys, siblings) have unique stimulus control over certain family behaviors, and therefore trigger certain reactions in children and parents that are absent in the clinic, limiting the generalizability of clinic-based observations. As evidenced by previous research (e.g., Webster-Stratton, 1985), clinical observations may elicit the predominant pattern of behaviors exhibited by parents and children, but may not capture the true magnitude of these behaviors as they are evoked at home. Thus, home-based observations may have incremental utility in certain circumstances where clinic-based observations are believed to have limited utility due to concerns with generalizability.

Although home-based observations may hold certain advantages in terms of external validity over clinic-based observations, there are also potential limitations to these types of observations that warrant consideration. First, it quickly became apparent during the gathering of home-based observational data for this project that there are numerous barriers to conducting standardized observations in the home. Apart from the logistical barriers (e.g., travel, safety, transporting equipment), there are a multitude of distractions that can be effectively eliminated during in-clinic observations that can jeopardize the “standardized” status and reliability of home-based observations. Televisions, phones, pets, siblings, loud appliances, and so on not only distract the family but also the observers collecting the data and the coders reviewing the data in the lab. These can cause errors in successfully gathering a standardized observation with fidelity and coding the observation reliably. Alternatively, these barriers and the novelty of conducting observations in a family’s home may force observers and coders to attend more to the observation procedure. Indeed, ICCs of coder reliability in this study were higher for in-home as compared to in-clinic observations, with 21 of the 22 DPICS codes being coded reliably in the home and 16 of the 22 being reliably coded in the clinic. Future studies should examine coder reliability as a function of observation stimulus value. For example, observations of “exciting” families (e.g., high noncompliance, high negative parent behaviors) could be compared with “boring” families (e.g., quiet, inactive). Furthermore, studies could examine coder reliability during in-home coding to determine if live coding is more or less reliable than in-lab coding of in-home DPICS recordings. The effect of observation setting and other factors on DPICS coder reliability are empirical questions that require further study; such studies could help inform the training of DPICS coders involved in research and clinicians completing certification in PCIT.

This study also revisits the topic of reactivity as it relates to analog observations of parent and child behavior. Although reactivity has been discussed for decades, there is still no general consensus for how to adequately measure and control for its effects. This study represents the first to measure reactivity, defined as a change in parent or child behavior as a result of being observed, using a parent-report measure. Although this study presents pilot data and is not sufficient to establish a psychometrically strong measure as ready for wide-spread clinical use, it does demonstrate that it is possible to measure reactivity using a parent report modality. This method would be substantially easier to use in controlling for reactivity during behavior observations than using previously-recommended methods, such as multiple observations to allow families to habituate to the observation procedures. Thus, measures of reactivity like the TORQ could be used to increase clinician confidence in the external validity of ABOs like the DPICS while maximizing the clinical utility and feasibility of using these observations. This study is also the first to consider parent perceived difficulty and comfort as important factors contributing to reactivity. Indeed, these variables were found to successfully predict parent and child behaviors during multiple segments of the DPICS. Also, the various subscales of the TORQ did not completely overlap as not all subscales significantly correlated with one another. Thus, it is believed that the various scales of the TORQ measure unique facets of reactivity, including child and parent behavioral representativeness, comfort, and difficulty. However, future editions of the TORQ may be improved by directly measuring obtrusiveness or incorporating the number of observers into the measure. Also, the TORQ allowed this study to measure reactivity across DPICS segments. The TORQ was also useful in predicting variance in parent and child behaviors observed during the DPICS. Thus, our study demonstrates that reactivity can play a role in behavioral observations, a role that can vary in magnitude with the

unique task demands of different analog situations. This role of reactivity can even be seen in non-clinical, community families and can no longer be ignored if ABOs are to be successfully integrated into EBA procedures.

Previous researchers have discussed potential variables contributing to reactivity and methods believed to minimize its effects (Aspland & Gardner, 2003; Gardner, 2000; Kazdin, 1982). The DPICS incorporates many of these suggestions in its protocol, including the use of warm-up segments, a one-way mirror, and a bug-in-ear device, to allow families time to acclimate to the observation tasks and to minimize obtrusiveness. However, previous studies show that some of these efforts may not be needed (Shanley & Niec, 2011; Thornberry & Brestan-Knight, 2011), and this study extends such findings by calling into question whether the number of observers present during observations is important when considering reactivity. In this study, number of observers did not significantly predict variance in TORQ scales but did predict variance in parent and child behaviors along with the TORQ. This finding suggests that researchers and clinicians can utilize multiple observers (up to three based on this study) or co-therapists without concern of significantly influencing reactivity, but doing so may influence the parent-child interaction. Future studies can use the TORQ to quantify the effects of the aforementioned strategies to decrease reactivity.

This study expanded on previous efforts to examine reactivity in that the TORQ included measures of parent comfort and perceived difficulty of observation instructions. These factors were important to measure in our conceptualization of parent and child reactivity. Nonsignificant correlations between these subscales of the TORQ suggest that each measure a unique proportion of variance in reactivity experienced during DPICS observations. In other words, this study highlights the importance of considering parent and child reactivity separately during behavior

observations. Further, examination of the relationships between various subscales of the TORQ provides insight into how parents reflect on their interactions with their children. For example, reactivity scores during the CLP and CU segments of the DPICS were positively correlated with child reactivity scores. This makes intuitive sense given that the child is supposed to be in the lead during CLP but not PLP, and a heavy demand is placed on the child during the CU segment of the DPICS. In contrast, parent reactivity factored less in CLP reactivity and factored more into PLP and CU segments, when parents are in more control of the interaction. Also, the nonsignificant positive correlation between parent and child reactivity scores suggests that these components of reactivity are not entirely independent; indeed, it can be easily inferred that a transactional relationship occurs in that parent reactivity begets child reactivity and vice versa. However, these data show that child and parent influence the interaction independently and directly, as well, and both should be considered separately in a comprehensive conceptualization of reactivity. Interestingly, parent perceived difficulty of the DPICS observation was highly correlated with segments in which parent reactivity was highest (PLP and CU) and was not significantly related to child reactivity. Future studies should analyze the relationships between these components of reactivity with clinical samples to determine if our conceptualization of reactivity is constant across populations or if there are systematic differences in parent reporting of reactivity across populations. Furthermore, adding child-reported reactivity measures may provide a more comprehensive picture of reactivity during dyadic interactions.

Measures like the TORQ can also serve a clinical purpose during parent-training interventions like PCIT. The TORQ could be used during initial DPICS observations to inform clinicians of the perceived representativeness of a given observation. Future studies could establish reactivity norms in clinical settings, which could be used to indicate above-average

levels of parent and/or child reactivity experienced during observations. Furthermore, and particularly with families with children that have severely disruptive behavior, it's possible that the TORQ could help parents become more aware of their child's behavior in the home. Parents may also become more aware of their own behavior as a result of responding to the TORQ, which could improve the magnitude and/or rate of parent behavior change during treatment. This improved awareness could also improve parents' implementation of strategies learned in treatment and improve the accuracy of their reports of practicing at home techniques learned in session. Indirectly, other parent-report measures of child behavior, such as the ECBI, can be used for the purpose of improving parent awareness of their and their child's behavior if the clinician chooses to review such measures with the parent periodically during treatment. The TORQ may allow clinicians to more directly address issues related to representativeness of parent and child behavior, parent perceived difficulty, and parent comfort during clinic-based observations. Future studies are needed to determine if the TORQ leads to improved outcomes in clinical treatment and improved parent awareness and reporting of behavior.

The findings of this study should be interpreted cautiously given several limitations that warrant consideration. First, our sample size is very small and represents a convenience sample of local community families that were not randomly assigned to groups. Although our two independent samples used for clinic- and home-based observations were very similar, they did differ on important demographic variables. In particular, there were more single and divorced parents in our home-based observation sample. Also, children in the clinic-based observation sample were rated to have significantly more aggressive behaviors than those observed in the home on the BASC-2. However, the average scores of both samples on all clinical scales administered fell within the normal range, suggesting that the impact of this difference is



minimal. Still, it is possible that unknown differences between the two observation setting groups could contribute to the few behavioral differences observed in this study. Future studies should seek to conduct observations in multiple settings using the same families in a counterbalanced manner such that these potential confounding variables can be controlled.

Our small sample also limits statistical power to detect differences between these two samples on the various measures used. Future studies with larger samples could provide more information regarding the influence of reactivity on parent and child behavior. For example, this study only analyzed the ability of TORQ scales to predict composite DPICS categories in order to limit family-wise Type I error. However, a future study with larger samples could analyze the impact of TORQ-measured reactivity on each of the 22 DPICS codes using more sophisticated statistical techniques to maximize power and minimize Type I error.

Another limitation of this study relates to the passive wording used in the TORQ. It was brought to our attention that such wording could be problematic when used with families whose primary language is not English. However, because data collection was ongoing at the time of this suggestion, it was decided not to incorporate this change into the present study in order to maximize the number of families' data used. Focus groups with a variety of families could be held in the future to revise the TORQ in order to improve comprehension and ease of administration.

This study was also limited by the low base rates of occurrence of some behaviors in this normative sample. Specifically, inappropriate parent and child behaviors were rare, as were some specialized parenting behaviors often targeted during parent-training interventions (e.g., Labeled Praise, Reflections). As a result, these rarely-occurring codes were not coded reliably (i.e.,  $\geq .75$  ICC). It is not surprising that these behaviors did not occur consistently in these samples given

the lack of clinically-significant behavioral problems in sample children and the lack of formal training in play therapy skills in sample parents. Thus, future studies with clinical populations or populations who have completed parent-training interventions who may display higher frequencies of these behaviors may produce higher DPICS coding reliabilities and different results related to the TORQ's ability to predict these behaviors. Furthermore, it should be noted that this normative sample did display some inappropriate behaviors (e.g., parent and child negative talk) and displayed a lack of certain appropriate parent behaviors (e.g., labeled praise, reflections, behavior descriptions, use of effective commands). These findings suggest the need for primary prevention efforts aimed at improving effective parenting practices in community families in order to provide psychoeducation to families and to prevent the development of coercive parent-child behavioral patterns.

Another limitation of this study is that we did not assess child perceptions of reactivity, that is, the child's perception of the representativeness of their or their parent's behavior, the child's comfort, or the child's perceived difficulty to follow task demands. Future studies should examine child-reported reactivity and compare these to parent reports on measures like the TORQ. Also, future studies could explore child comfort levels and directly ask children how comfortable they are during home- and clinic-based observations across various observation tasks. This could help clinicians anticipate which portions of treatment may be more intimidating for a child and could help them prepare parents for sessions in which uncomfortable children may be more prone to act out. Given the low number of items on the TORQ, it may be psychometrically beneficial to expand the TORQ to include child-completed items to increase total variance collected related to parent and child reactivity.

In sum, this project represents a next step in the growing efforts of the field to develop EBAs to accompany the progress of EBTs. In particular, this study bolsters the external validity of the DPICS, an ABO of parent-child interactions, by documenting home-based normative data and comparing these data to normative data gathered in a clinic setting. This study is also the first to document behavioral norms for nonclinical families using the abridged DPICS. Extending the psychometric support of ABOs like the DPICS to the home setting in an efficient manner is an important prerequisite for developing home-based EBTs. Furthermore, this study presents a new, parent-report assessment modality for measuring reactivity, a validity threat to behavior observations that can be difficult and time-consuming to minimize. It is hoped that this measure of reactivity will improve the external validity of behavior observations without hindering clinical utility, allowing behavioral observations to gain widespread use and incorporation into developing EBA guidelines.

## References

- Abidin, R. (1995). *Parenting stress index*. Psychological Assessment Resources, Inc. (3<sup>rd</sup> ed.). Florida: Odessa.
- Achenbach, T. (2005). Advancing assessment of children and adolescents: Commentary on evidence-based assessment of child and adolescent disorders. *Journal of Clinical Child and Adolescent Psychology*, *34*, 541-547.
- Achenbach, T., & Rescorla, L. (2001). *Manual for the ASEBA School-Age forms & profiles*. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.
- American Psychological Association Presidential Task Force on Evidence-Based Practice (2006).
- Aspland, H., & Gardner, F. (2003). Observational measures of parent-child interaction: An introductory review. *Child and Adolescent Mental Health*, *8*, 136-143.
- Bagner, D., & Eyberg, S. (2007). Parent-Child Interaction Therapy for disruptive behavior in children with mental retardation: A randomized controlled trial. *Journal of Clinical Child and Adolescent Psychology*, *36*, 418-429.
- Bagner, D., Fernandez, M., & Eyberg, S. (2004). Parent-Child Interaction Therapy and chronic illness: A case study. *Journal of Clinical Psychology in Medical Settings*, *11*, 1-6.

- Basco, M., Bostic, J., Davies, D., Rush, A., Witte, B., Hendrickse, W., et al. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *American Journal of Psychiatry, 157*, 1599-1605.
- Baum, C., Forehand, R., & Zegiob, L. (1979). A review of observer reactivity in adult-child interactions. *Journal of Behavioral Assessment, 1*, 167-178.
- Bernal, M., Gibson, D., William, D., & Pesses, D. (1971). A device for automatic audio tape recording. *Journal of Applied Behavior Analysis, 4*, 151-156.
- Bessmer, J. (1998). *The Dyadic Parent-Child Interaction Coding System II (DPICS II): Reliability and validity with mother-child dyads*. Unpublished doctoral dissertation, Auburn University, Auburn, Alabama.
- Bessmer, J., & Eyberg, S. (1993, November). *Dyadic Parent-Child Interaction Coding System II (DPICS II): Initial reliability and validity of the clinical version*. Paper presented at the AABT Preconference on Social Learning and the Family, Atlanta, GA.
- Brestan, E., Eyberg, S., Boggs, S., & Algina, J. (1997). Parent-Child Interaction Therapy: Parents' perceptions of untreated siblings. *Child & Family Behavior Therapy, 19*, 13-28.
- Brestan-Knight, E., & Salamone, C.A. (2011). Measuring parent-child interactions through play. In S. Russ & L. Niec (Eds.), *Play in clinical practice: Evidence-based approaches* (pp. 83-108). New York: Guilford Publications, Inc.
- Brinkmeyer, M. (2006). *Conduct disorder in young children: A comparison of clinical presentation and treatment outcome in preschoolers with conduct disorder versus oppositional defiant disorder*. Unpublished doctoral dissertation, University of Florida, Gainesville, Florida.

- Cantor, D., & Fuentes, M. (2008). Psychology's response to managed care. *Professional Psychology: Research and Practice, 39*, 638-645.
- Cashel, M. (2002). Child and adolescent psychological assessment: Current clinical practices and the impact of managed care. *Professional Psychology: Research and Practice, 33*, 446-453.
- Chaffin, M., Silovsky, J., Funderburk, B., Valle, L., Brestan, E., Balachova, T., Jackson, S., Lensgraf, J., & Bonner, B. (2004). Parent-Child Interaction Therapy with physically abusive parents: Efficacy for reducing future abuse reports. *Journal of Consulting and Clinical Psychology, 72*, 500-510.
- Chambless, D., Baker, M., Baucom, D., Beutler, L., Calhoun, K., Crits-Christoph, P., et al. (1998). Update on empirically validated therapies: II. *The Clinical Psychologist, 51*, 3-16.
- Chambless, D., Sanderson, W., Shoham, V., Bennet Johnson, S., Pope, K., Crits-Christoph, P., et al. (1996). An update on empirically validated therapies. *The Clinical Psychologist, 49*, 5-18.
- Chase, R., & Eyberg, S. (2006). *Abridged Manual for the Dyadic Parent-Child Interaction Coding System (3<sup>rd</sup> edition)*. Retrieved May 13, 2010, from the Parent-Child Interaction Therapy website: <http://www.pcit.org>.
- Choate, M., Pincus, D., Eyberg, S., & Barlow, D. (2005). Parent-Child Interaction Therapy for treatment of separation anxiety disorder in young children: A pilot study. *Cognitive and Behavioral Practice, 12*, 126-135.
- Cohen, L., La Greca, A., Blount, R., Kazak, A., Holmbeck, G., & Lemanek, K. (2008). Introduction to special issue: Evidence-based assessment in pediatric psychology. *Journal of Pediatric Psychology, 33*, 911-915.

- Coursen, L. (2009). *Frequencies of DPICS-III codes for a sample of 8 to 12 year olds*. Unpublished honor's thesis, Auburn University, Auburn, Alabama.
- Deskins, M. (2005). *The Dyadic Parent-Child Interaction Coding System II: Reliability and validity with school-aged dyads*. Unpublished doctoral dissertation, Auburn University, Auburn, Alabama.
- Dunn, J., & Kendrick, C. (1980). The arrival of a sibling: Changes in patterns of interaction between mother and first-born child. *Journal of Child Psychology and Psychiatry, 21*, 119-132.
- Dunn, J., & Kendrick, C. (1982). *Siblings: Love, envy and understanding*. London: Grant McIntyre.
- Eddy, J., Dishion, T., & Stoolmiller, M. (1998). The analysis of intervention change in children and families: Methodological and conceptual issues embedded in intervention studies. *Journal of Abnormal Child Psychology, 26*, 53-71.
- Eisman, E., Dies, R., Finn, S., Eyde, L., Kay, G., Kubiszyn, T., et al. (2000). Problems and limitations in using psychological assessment in the contemporary health care delivery system. *Professional Psychology: Research and Practice, 31*, 131-140.
- Eyberg, S. (2010). *Parent-Child Interaction Therapy: Integrity Checklists and Session Materials*. Unpublished treatment manual.
- Eyberg, S., Nelson, M., Duke, M., & Boggs, S. (2005). *Manual for the Dyadic Parent-Child Interaction Coding System* (3rd ed.). Retrieved July 28, 2006 from The University of Florida, Parent-Child Interaction Therapy Web site: [www.pcit.org](http://www.pcit.org).

- Eyberg, S., & Pincus, D. (1999). *Eyberg Child Behavior Inventory and Sutter-Eyberg Student Behavior Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Fagot, B. & Leve, L. (1998). Teacher ratings of externalizing behavior at school entry for boys and girls: Similar early predictors and different correlates. *Journal of Child Psychology and Psychiatry*, 39, 555-566.
- Fergusson, D., Lynskey, M., & Horwood, L. (1993). The effect of maternal depression on maternal ratings of child behavior. *Journal of Abnormal Child Psychology*, 21, 245-271.
- Foote, R. (2000). *The Dyadic Parent-Child Interaction Coding System II (DPICS II): Reliability and validity with father-child dyads*. Unpublished doctoral dissertation, Auburn University, Auburn, Alabama.
- Forgatch, M., & DeGarmo, D. (1999). Parenting through change: An effective parenting training program for single mothers. *Journal of Consulting and Clinical Psychology*, 67, 711-724.
- Frick, P. & McMahon, R. (2008). Child and adolescent conduct problems. In J. Hunsley & E. Mash (Eds.), *A Guide to Assessments that Work* (pp. 41-66). New York: Oxford University.
- Funderburk, B., Eyberg, S., Newcomb, K., McNeil, C., Hembree-Kigin, T., & Capage, L. (1998). Parent-child interaction therapy with behavior problem children: Maintenance of treatment effects in the school setting. *Child & Family Behavior Therapy*, 20, 17-38.
- Galanter, R., Self-Brown, S., Valente, J., Dorsey, S., Whitaker, D., Bertuglia-Haley, M., et al. (2012). Effectiveness of Parent-Child Interaction Therapy delivered to at-risk families in the home setting. *Child & Family Behavior Therapy*, 34, 177-196.



- Gardner, F. (1987). Positive interaction between mothers and children with conduct problems: Is there training for harmony as well as fighting? *Journal of Abnormal Child Psychology*, *15*, 283-293.
- Gardner, F. (2000). Methodological issues in the direct observation of parent-child interaction: Do observational findings reflect the natural behavior of participants? *Clinical Child and Family Psychology Review*, *3*, 185-198.
- Gillaspy, S. & Bonner, B. (2010). Child maltreatment. In M. Roberts & R. Steele (Eds.), *Handbook of Pediatric Psychology (4<sup>th</sup> ed.)*. (pp. 556-571). New York: The Guilford Press.
- Hanf, C. (1969). *A two stage program for modifying maternal controlling during the mother-child interaction*. Paper presented at the meeting of the Western Psychological Association, Vancouver, British Columbia.
- Hartmann, D., & Wood, D. (1990). Observational methods. In A. Bellack, M. Hersen, & A. Kazdin, (Eds.), *International handbook of behavior modification and therapy* (pp. 107-138). New York: Plenum.
- Haynes, S. (2001). Clinical applications of analogue behavioral observation: Dimensions of psychometric evaluation. *Psychological Assessment*, *13*, 73-85.
- Herbert, E. & Baer, D. (1972). Training parents as behavior modifiers: Self-recording of contingent attention. *Journal of Applied Behavior Analysis*, *5*, 139-149.
- Heyman, R., & Slep, A. (2004). Analogue behavioral observation. In S. Haynes and E. Heiby's (Eds.) *Comprehensive Handbook of Psychological Assessment, Vol. 3: Behavioral Assessment*. Hoboken, NJ, US: John Wiley & Sons, Inc.

- Hughes, M., Carmichael, H., Pinkerton, G., & Tizard, B. (1979). Recording children's conversations at home and at nursery school: A technique and some methodological considerations. *Journal of Child Psychology and Psychiatry*, 20, 225-232.
- Hunsley, J. & Mash, E. (2011). Evidence based assessment. In D. Barlow (Ed.), *Oxford Handbook of Clinical Psychology* (pp. 76-97). New York: Oxford University Press.
- Hunsley, J. & Mash, E. (2010). The role of assessment in evidence-based practice. In M. Antony & D. Barlow (Eds.), *Handbook of Assessment and Treatment Planning for Psychological Disorders (2<sup>nd</sup> ed.)* (pp. 3-22). New York: Guilford Press.
- Hunsley, J. & Mash, E. (2008a) (Eds.), *A Guide to Assessments that Work*. New York: Oxford University Press.
- Hunsley, J. & Mash, E. (2008b). Developing criteria for evidence-based assessment: An introduction to assessments that work. In J. Hunsley & E. Mash (Eds.), *A Guide to Assessments that Work*. (pp. 3-14). New York: Oxford University Press.
- Hunsley, J. & Mash, E. (2005). Introduction to the special section on developing guidelines for the evidence-based assessment of adult disorders. *Psychological Assessment*, 17, 251-255.
- Jacob, T., Tennenbaum, D., Bargiel, K., & Seilhamer, R. (1995). Family interaction in the home: Development of a new coding scheme. *Behavior Modification*, 19, 147-169.
- Jacob, T., Tennenbaum, D., Seilhamer, R., Bargiel, K., et al. (1994). Reactivity effects during naturalistic observation of distressed and nondistressed families. *Journal of Family Psychology*, 8, 354-363.

- Jensen-Doss, A. (2011). Practice involves more than treatment: How can evidence-based assessment catch up to evidence-based treatment? *Clinical Psychology Science and Practice, 18*, 173-177.
- Jensen-Doss, A. & Hawley, K. (2010). Understanding barriers to evidence-based assessment: Clinician attitudes toward standardized assessment tools. *Journal of Clinical Child and Adolescent Psychology, 39*, 885-896.
- Johnson, S., & Bolstad, O. (1975). Reactivity to home observation: A comparison of audio recorded behavior with observers present or absent. *Journal of Applied Behavior Analysis, 8*, 181-185.
- Johnson, S., & Lobitz, G. (1974). Parental manipulation of child behaviors in home observations. *Journal of Applied Behavior Analysis, 7*, 23-31.
- Kazdin, A. (1982). Observer effects: reactivity of direct observation. In D. Hartmann (Ed.), *Using observers to study behavior*. (pp. 5-19). San Francisco: Jossey Bass.
- Kier, C. (1996). How natural is “naturalistic” home observation? Observer reactivity in infant-sibling interaction. *Proceedings of the British Psychological Society, 4*, 79.
- Lewis, C., Kier, C., Hyder, C., Prenderville, N., Pullen, J., & Stephens, A. (1996). Observer influences on fathers and mothers: An experimental manipulation of the structure and function of parent-infant conversation. *Early Development and Parenting, 5*, 57-68.
- Luebke, A., Radcliffe, A., Callands, T., Green, D., & Thorn, B. (2007). Evidence-based practice in psychology: Perceptions of graduate students in scientist-practitioner programs. *Journal of Clinical Psychology, 63*, 643-655.

- Margolin, G., Oliver, P., Gordis, E., O'Hearn, H., Medina, A., Ghosh, C., & Morland, L. (1998). The nuts and bolts of behavioral observation of marital and family interaction. *Clinical Child and Family Psychology Review, 1*, 195-213.
- Mash, E., & Foster, S. (2001). Exporting analogue behavioral observation from research to clinical practice: Useful or cost-defective? *Psychological Assessment, 13*, 86-98.
- Mash, E., & Hunsley, J. (2005). Evidence-based assessment of child and adolescent disorders: Issues and challenges. *Journal of Clinical Child and Adolescent Psychology, 34*, 362-379.
- Masse, J. & McNeil, C. (2008). In-home Parent-Child Interaction Therapy: Clinical considerations. *Child & Family Behavior Therapy, 30*, 127-135.
- McMahon, R., & Frick, P. (2005). Evidence-based assessment of conduct problems in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*, 477-505.
- McNeil, C., Eyberg, S., Eisenstadt, T., Newcomb, K., & Funderburk, B. (1991). Parent-child interaction therapy with behavior problem children: Generalization of treatment effects to the school setting. *Journal of Clinical Child Psychology, 20*, 140-151.
- Miller, E., & Eyberg, S. (1991). Parent-child interaction therapy with a diabetic child. In S. Boggs & C. Rodriguez (Eds.), *Advances in Child Health Psychology: Abstracts*. Gainesville, FL: Clinical and Health Psychology Publishing.
- Mori, L., & Armendariz, G. (2001). Analogue assessment of child behavior problems. *Psychological Assessment, 13*, 36-45.
- Nelson, T. & Steele, R. (2008). Influences on practitioner treatment selection: Best research evidence and other considerations. *Journal of Behavioral Health Services and Research, 35*, 170-178.

- Nixon, R., Sweeney, L., Erickson, D., & Touyz, S. (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.
- Palmiter, D. (2004). A survey of the assessment practices of child and adolescent clinicians. *American Journal of Orthopsychiatry, 74*, 122-128.
- Patterson, G. (1982). *Coercive family process*. Eugene, OR: Castalia.
- Patterson, G., & Forgatch, M. (1995). Predicting future clinical adjustment from treatment outcome and process variables. Special Issue: Methodological issues in psychological assessment research. *Psychological Assessment, 7*, 275-285.
- Pepler, D., & Craig, W. (1995). A peek behind the fence: Naturalistic observations of aggressive children with remote audio-visual recording. *Developmental Psychology, 31*, 548-553.
- Pett, M., Wampold, B., Baughan-Cole, B., & East, T. (1992). Consistency of behaviors within a naturalistic setting: An examination of the impact of context and repeated observations on mother-child interactions. *Behavioral Assessment, 14*, 367-385.
- Piotrowski, C., Belter, R., & Keller, J. (1998). The impact of managed care on the practice of psychological testing: Preliminary findings. *Journal of Personality Assessment, 70*, 441-447.
- Prescott, A., Bank, L., Reid, J., Knutson, J., Burraston, B., & Eddy, J. (2000). The veridicality of punitive childhood experiences reported by adolescents and young adults. *Child Abuse and Neglect, 24*, 411-425.
- Rapoport, J., & Benoit, M. (1975). The relation of direct home observations to the clinic evaluation of hyperactive school age boys. *Journal of Child Psychology and Psychiatry, 16*, 141-147.

- Rettew, D., Lynch, A., Achenbach, T., Dumenci, L., & Ivanova, M. (2009). Meta-analyses of agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. *International Journal of Methods in Psychiatric Research, 18*, 169-184.
- Reynolds, C. & Kamphaus, R. (2004). *Behavior Assessment System for Children* (2nd ed.). Circle Pines, MN: American Guidance Services.
- Rhule, D., McMahon, R., & Vando, J. (2009). The acceptability and representativeness of standardized parent-child interaction tasks. *Behavior Therapy, 40*, 393-402.
- Richters, J. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin, 112*, 485-499.
- Roberts, M. (2001). Clinic observations of structured parent-child interaction designed to evaluate externalizing disorders. *Psychological Assessment, 13*, 46-58.
- Roberts, M., & Powers, S. (1988). The Compliance Test. *Behavioral Assessment, 10*, 375-398.
- Robinson, E., & Eyberg, S. (1981). The dyadic parent-child interaction coding system: Standardization and validation. *Journal of Consulting & Clinical Psychology, 49*(2), 245-250.
- Russell, A., Russell, G., & Midwinter, D. (1992). Observer influences on mothers and fathers: Self reported influence during a home observation. *Merrill-Palmer Quarterly, 36*, 263-283.
- Sanders, M., Markie-Dadds, C., Tully, L., & Bor, W. (2000). The Triple P-Positive Parenting Program: A comparison of enhanced, standard and self-directed behavioral family intervention for parents of children with early onset conduct problems. *Journal of Consulting and Clinical Psychology, 68*, 624-640.

- Schuhmann, E., Foote, R., Eyberg, S., Boggs, S., & Algina, J. (1998). Parent-Child Interaction Therapy: Interim report of a randomized trial with short-term maintenance. *Journal of Clinical Child Psychology, 27*, 34-45.
- Shanley, J. & Niec, L. (2011). The contribution of the Dyadic Parent-Child Interaction Coding System (DPICS) warm-up segments in assessing parent-child interactions. *Child & Family Behavior Therapy, 33*, 248-263.
- Stout, C. & Cook, L. (1999). New areas for psychological assessment in general health care settings: What to do today to prepare for tomorrow. *Journal of Clinical Psychology, 55*, 797-812.
- Thornberry, Jr., T., & Brestan-Knight, E. (2011). Analyzing the utility of Dyadic Parent-Child Interaction Coding System (DPICS) warm-up segments. *Journal of Psychopathology and Behavioral Assessment, 33*, 187-195.
- Timmer, S., Zebell, N., Culver, M., & Urquiza, A. (2009). Efficacy of adjunct in-home coaching to improve outcomes in Parent-Child Interaction Therapy. *Research on Social Work Practice, 20*, 36-45.
- Turchik, J., Karpenko, V., Hammers, D., & McNamara, J. (2007). Practical and ethical assessment issues in rural, impoverished, and managed care settings. *Professional Psychology: Research and Practice, 38*, 158-168.
- Ware, L., Fortson, B., & McNeil, C. (2003). Parent-Child Interaction Therapy: A promising intervention for abusive families. *The Behavior Analyst Today, 3*, 375-382.
- Ware, L., McNeil, C., Masse, J., & Stevens, S. (2008). Efficacy of in-home Parent-Child Interaction Therapy. *Child & Family Behavior Therapy, 30*, 99-126.

- Webster-Stratton, C. (1985). Comparisons of behavior transactions between conduct-disordered children and their mothers in the clinic and at home. *Journal of Abnormal Child Psychology, 13*, 169-184.
- Webster-Stratton, C. (1994). Advancing videotape parent training: A comparison study. *Journal of Consulting and Clinical Psychology, 62*, 583-593.
- Webster-Stratton, C. (1998). Preventing conduct problems in head start children: Strengthening parenting competencies. *Journal of Consulting and Clinical Psychology, 66*, 715-730.
- White, G. (1977). The effects of observer presence on the activity level of families. *Journal of Applied Behavior Analysis, 10*, 734.
- Wilsie, C., Travis, J., Thornberry, Jr., T., & Brestan-Knight, E. (October, 2010). *Evaluating trainee outcomes following a 40-hour face-to-face PCIT training*. Poster presented at the 2010 Kansas Conference in Clinical Child and Adolescent Psychology, Lawrence, Kansas.
- Zegiob, L. & Forehand, R. (1978). Parent-child interactions: Observer effects and social class differences. *Behavior Therapy, 9*, 118-123.



## Appendix

### Demographic Questionnaire

On this page, please provide information about **YOURSELF**

**Your Age** \_\_\_\_\_

**Relationship to child** \_\_\_\_\_

(mother, father, grandparent, relative, guardian, etc)

**Gender** \_\_\_\_\_

**Other Caregiver's Relationship to child** \_\_\_\_\_

**Ethnicity (please pick the best one that identifies your cultural group)**

- \_\_\_\_\_ African American
- \_\_\_\_\_ Asian (born in an Asian country)
- \_\_\_\_\_ Asian American (born in the United States)
- \_\_\_\_\_ Caucasian
- \_\_\_\_\_ Hispanic (non-white)
- \_\_\_\_\_ Multicultural
- \_\_\_\_\_ Native American
- \_\_\_\_\_ Other (please specify)

**Marital Status**

- \_\_\_\_\_ Married
- \_\_\_\_\_ Divorced
- \_\_\_\_\_ Remarried
- \_\_\_\_\_ Widowed
- \_\_\_\_\_ Single

**Highest level of education completed**

(check one for each if applicable)

Yourself	Spouse
<input type="checkbox"/>	<input type="checkbox"/>
9th grade	
<input type="checkbox"/>	<input type="checkbox"/>
10 years	

**Approximate total household income**

(please include yourself and others in the home)

Please check one

**Please check one**

- \_\_\_ less than \$10,000
- \_\_\_ \$10,000-\$15,000

		11 years	___ \$15,000-\$20,000
		12 years	___ \$20,000-\$25,000
		Some college (1-2 years)	___ \$25,000-\$30,000
		Associates degree	___ \$30,000-40,000
		Bachelors degree	___ \$40,000-\$50,000
		Masters degree	___ \$50,000-60,000
		Doctoral degree	___ \$60,000-\$70,000
			___ \$70,000-\$80,000
			___ \$80,000-\$90,000
			___ \$90,000-\$100,000
			___ more than \$100,000

**Current Occupation** \_\_\_\_\_

**Spouse Occupation** \_\_\_\_\_

**Current Field of Work** \_\_\_\_\_

**Spouse Field of Work** \_\_\_\_\_

### Demographic Questionnaire

On this page, please provide information about your **CHILD**

**Please fill out the following information in regards to your child who is participating**

**Date of Birth** \_\_\_\_\_

**Age** \_\_\_\_\_

**Gender** \_\_\_\_\_

**Ethnicity** (please pick the best answer that identifies your child's cultural group)

\_\_\_\_\_ African American

\_\_\_\_\_ Asian (born in an Asian country)

\_\_\_\_\_ Asian American (born in the United States)

\_\_\_\_\_ Caucasian

\_\_\_\_\_ Hispanic (non-white)

\_\_\_\_\_ Multicultural

\_\_\_\_\_ Native American

\_\_\_\_\_ Other (please specify)

**Please fill out the following information regarding other children in the home**

Number of children in the home between 2 and 10 \_\_\_\_\_

TORQ

Regarding your play time with your child today, please answer the following questions by circling the appropriate response.

**When you were letting YOUR CHILD lead the play . . .**

1. How did YOUR CHILD'S behavior compare to his/her typical behavior?

1 Very Different	2 Somewhat Different	3 Somewhat Similar	4 Very Similar
------------------------	----------------------------	--------------------------	----------------------

How was his/her behavior similar/different?

---

2. How did YOUR behavior compare to your typical behavior?

1 Very Different	2 Somewhat Different	3 Somewhat Similar	4 Very Similar
------------------------	----------------------------	--------------------------	----------------------

How was your behavior similar/different?

---

3. How difficult was it to follow the given directions when you were asked to let your child lead the play?

1 Very Difficult	2 Somewhat Difficult	3 Somewhat Easy	4 Very Easy
------------------------	----------------------------	-----------------------	-------------------

How was it easy/difficult?

---

4. How comfortable did you feel allowing your child to lead the play?

1 Very Comfortable	2 Somewhat Comfortable	3 Somewhat Uncomfortable	4 Very Uncomfortable
--------------------------	------------------------------	--------------------------------	----------------------------

**When YOU were leading the play . . .**

5. How did YOUR CHILD'S behavior compare to his/her typical behavior?

1 Very Different	2 Somewhat Different	3 Somewhat Similar	4 Very Similar
------------------------	----------------------------	--------------------------	----------------------

How was his/her behavior similar/different?

---

6. How did YOUR behavior compare to your typical behavior?

1 Very Different	2 Somewhat Different	3 Somewhat Similar	4 Very Similar
------------------------	----------------------------	--------------------------	----------------------

How was your behavior similar/different?

---

7. How difficult was it to follow the given directions when you were asked to lead the play?

1 Very Difficult	2 Somewhat Difficult	3 Somewhat Easy	4 Very Easy
------------------------	----------------------------	-----------------------	-------------------

How was it easy/difficult?

---

8. How comfortable did you feel leading the play?

1 Very Comfortable	2 Somewhat Comfortable	3 Somewhat Uncomfortable	4 Very Uncomfortable
--------------------------	------------------------------	--------------------------------	----------------------------

**When you asked your child to clean up all the toys . . .**

9. How did YOUR CHILD'S behavior compare to his/her typical behavior?

1 Very Different	2 Somewhat Different	3 Somewhat Similar	4 Very Similar
------------------------	----------------------------	--------------------------	----------------------

How was his/her behavior similar/different?

---

10. How did YOUR behavior compare to your typical behavior?

1 Very Different	2 Somewhat Different	3 Somewhat Similar	4 Very Similar
------------------------	----------------------------	--------------------------	----------------------

How was your behavior similar/different?

---

11. How difficult was it to follow the given directions when you were asked to have your child clean up?

1 Very Difficult	2 Somewhat Difficult	3 Somewhat Easy	4 Very Easy
------------------------	----------------------------	-----------------------	-------------------

How was it easy/difficult?

---

12. How comfortable did you feel asking your child to clean up?

1 Very Comfortable	2 Somewhat Comfortable	3 Somewhat Uncomfortable	4 Very Uncomfortable
--------------------------	------------------------------	--------------------------------	----------------------------

**Overall . . .**

13. How did YOUR CHILD'S behavior compare to his/her typical behavior?

1 Very Different	2 Somewhat Different	3 Somewhat Similar	4 Very Similar
------------------------	----------------------------	--------------------------	----------------------

How was his/her behavior similar/different?

---

14. How did YOUR behavior compare to your typical behavior?

1 Very Different	2 Somewhat Different	3 Somewhat Similar	4 Very Similar
------------------------	----------------------------	--------------------------	----------------------

How was your behavior similar/different?

---

15. How difficult was it to follow the given directions during the play?

1 Very Difficult	2 Somewhat Difficult	3 Somewhat Easy	4 Very Easy
------------------------	----------------------------	-----------------------	-------------------

How was it easy/difficult?

---

16. How comfortable did you feel during the play?

1 Very Comfortable	2 Somewhat Comfortable	3 Somewhat Uncomfortable	4 Very Uncomfortable
--------------------------	------------------------------	--------------------------------	----------------------------

17. Please provide any other comments related to your thoughts/feelings of being observed today:

---

---

# In-Home DPICS Parent Coding Sheet

Tape #: \_\_\_\_\_ Coders: \_\_\_\_\_

Circle DPICS Segment: WCLP CLP WPLP PLP CU

Circle One: Primary Reliability Segment Start Time: \_\_\_\_\_

Behavior	Count			Total
<b>TA</b>				
<b>BD</b>				
<b>RF</b>				
<b>UP</b>				
<b>LP</b>				
<b>NTA</b>				
<b>DQ</b>				
	<b>AN/CO</b>	<b>NA/NC</b>	<b>NOA/NOC</b>	
<b>IQ</b>				
<b>DC</b>				
<b>IC</b>				
<b>PTO</b>				
<b>NTO</b>				



# In-Home DPICS Child Coding Sheet

Tape #: \_\_\_\_\_

Coder: \_\_\_\_\_

Circle DPICS Segment:    WCLP    CLP    WPLP    PLP    CU

Circle One:    Primary

Reliability

Segment Start Time: \_\_\_\_\_

Behavior	Count	Total
<b>PRO</b>		
<b>QU</b>		
<b>CM</b>		
<b>NTA</b>		
<b>YE</b>		
<b>WH</b>		
<b>PTO</b>		
<b>NTO</b>		

Table 1

*Parent and Child Behaviors and Respective DPICS-III Codes*

Parent Behavior (Code)	Child Behavior (Code)
Negative Talk (NTA)	Negative Talk (NTA)
Direct Command (DC)	Command (CM)
Indirect Command (IC)	Question (QU)
Labeled Praise (LP)	Prosocial Talk (PRO)
Unlabeled Praise (UP)	Yell (YE)
Information Question (IQ)	Whine (WH)
Descriptive/Reflective Question (DQ)	Answer (AN)
Reflective Statement (RF)	No Answer (NA)
Behavioral Description (BD)	No Opportunity for Answer (NOA)
Neutral Talk (TA)	Comply (CO)
Negative Touch (NTO)	Noncomply (NC)
Positive Touch (PTO)	No Opportunity for Compliance (NOC)
	Negative Touch (NTO)
	Positive Touch (PTO)

Table 2

*DPICS-III Parent and Child Composite Categories and Respective Formulae*

Composite Category	Formula
Parent Inappropriate Behavior	$pIQ + pDQ + pNTA$
Parent Prosocial Behavior	$pBD + pRF + pUP + pLP$
Child Compliance	$cCO \div [(pDC + pIC) - cNOC]$
Child Noncompliance	$cNC \div [(pDC + pIC) - cNOC]$
Child Inappropriate Behavior	$cNTA + cYE + cWH$

Note: The subscripts  $c$  and  $p$  denote child and parent categories, respectively. Adapted from Eyberg et al. (2005).

Table 3

*Demographic and Behavior Rating Information by Observation Setting (N=27)*

<u>Variable</u>	<u>Clinic-Based (n=15)</u>	<u>Home-Based (n=12)</u>	<u>Ind. Samples <i>t</i></u>	<u>df</u>	<u><i>p</i> (two-tailed)</u>
<i>Continuous Variables</i>		Mean (SD)			
Parent age	35.00 (5.56)	31.33 (7.60)	1.45	25	.16
Child age	4.40 (1.55)	4.25 (2.01)	0.22	25	.83
Children in the home	1.33 (1.05)	1.33 (0.65)	0.00	25	1.00
<i>Categorical Variables</i>		Frequency (%)			
Parent gender:			0.16	25	.88
Biological mother	14 (93.3%)	11 (91.7%)			
Biological father	1 (6.7%)	1 (8.3%)			
Parent ethnicity:			1.95	25	.06
African American	2 (13.3%)	5 (41.7%)			
Caucasian	11 (73.3%)	7 (58.3%)			
Other	2 (13.4%)	0			
*Marital status:			2.18	25	.04
Married	14 (93.3%)	7 (58.3%)			

Divorced	0	2 (16.7%)			
Remarried	1 (6.7%)	0			
Single	0	3 (25%)			
Parent education:			0.68	25	.50
Some college	2 (13.3%)	4 (33.3%)			
Associate's	0	1 (8.3%)			
Bachelor's	5 (33.3%)	1 (8.3%)			
Master's	8 (53.3%)	4 (33.3%)			
Doctoral	0	2 (16.7%)			
Spouse education:			1.10	25	.28
High school	0	1 (8.3%)			
Some college	2 (13.3%)	1 (8.3%)			
Bachelor's	3 (20.0%)	4 (33.3%)			
Master's	6 (40.0%)	1 (8.3%)			
Doctoral	4 (26.7%)	2 (16.7%)			
Missing	0	3 (25.0%)			
Family annual income:			0.66	25	.51

\$0 – 10,000	0	2 (16.7%)		
\$10,000 – 15,000	0	1 (8.3%)		
\$20,000 – 25,000	2 (13.3%)	2 (16.7%)		
\$25,000 – 30,000	1 (6.7%)	0		
\$50,000 – 60,000	2 (13.3%)	0		
\$60,000 – 70,000	4 (26.7%)	0		
\$70,000 – 80,000	1 (6.7%)	2 (16.7%)		
\$80,000 – 90,000	2 (13.3%)	1 (8.3%)		
\$90,000 – 100,000	2 (13.3%)	0		
>\$100,000	1 (6.7%)	4 (33.3%)		
Child gender:			0.17	25 .87
Female	7 (46.7%)	6 (50%)		
Male	8 (53.3%)	6 (50%)		
Child ethnicity:			1.95	25 .06
African American	2 (13.3%)	5 (41.7%)		
Caucasian	11 (73.3%)	7 (58.3%)		
Other	2 (13.4%)	0		

---

$*p < .05$

Table 4

*BASC-2 and ECBI Scores by Observation Setting*

Scale	Mean (SD)		<i>F</i>	Cohen's <i>d</i>
	Clinic (n=15)	Home (n=12)		
<u><i>BASC-2 Composites (T scores)</i></u>				
Externalizing	49.80 (7.36)	44.58 (7.34)	3.36	0.74
Internalizing	48.73 (5.99)	47.00 (9.11)	0.35	0.24
BSI	48.13 (7.70)	45.50 (5.90)	0.95	0.39
Adapt. Skills	54.60 (7.81)	52.33 (7.75)	0.57	0.30
<u><i>BASC-2 Subscales (T scores)</i></u>				
Hyperactivity	51.00 (7.29)	48.75 (8.14)	0.57	0.30
*Aggression	49.73 (9.22)	41.67 (6.84)	6.36	1.01
Anxiety	50.87 (7.16)	51.25 (10.19)	-0.01	-0.05
Depression	50.33 (7.40)	46.17 (7.37)	2.12	0.59
Somatization	46.07 (7.06)	45.42 (7.90)	0.05	0.09
Atypicality	45.13 (5.95)	47.58 (5.85)	-1.15	-0.43
Withdrawal	46.20 (9.14)	46.67 (6.67)	-0.02	-0.06
Attention Prob.	49.60 (10.68)	48.75 (7.64)	0.05	0.09
Adaptability	54.27 (9.63)	50.58 (7.54)	1.18	0.44
Social Skills	57.20 (8.87)	53.67 (8.98)	1.05	0.41
Resiliency	54.27 (7.35)	52.33 (4.98)	0.61	0.31
<u><i>ECBI (Raw scores)</i></u>				
Intensity	99.40 (24.49)	89.50 (25.73)	1.04	0.41
Problem	8.40 (6.82)	6.42 (5.48)	0.67	0.33

\**p*<.05



Table 5

*TORQ Descriptive Statistics (N=27)*

Scale	Range	M (SD)	Skewness (SE)	Kurtosis (SE)
CLP Total	4 – 9	6.22 (1.85)	0.12 (0.45)	-1.38 (0.87)
PLP Total	4 – 13	8.00 (2.47)	0.02 (0.45)	-0.70 (0.87)
CU Total	4 – 13	7.37 (2.66)	0.31 (0.45)	-0.86 (0.87)
Total Reactivity	16 – 43	27.67 (7.20)	0.03 (0.45)	-0.71 (0.87)
Child Represent.	4 – 12	5.85 (2.11)	1.46 (0.45)	2.28 (0.87)
Parent Represent.	4 – 16	8.07 (3.06)	0.44 (0.45)	0.08 (0.87)
Difficulty	4 – 11	6.44 (2.10)	0.62 (0.45)	-0.72 (0.87)
Comfort	4 – 16	7.26 (3.47)	1.13 (0.45)	0.33 (0.87)

Table 6

*TORQ Correlations between Scales (N=27)*

Scale	CLP	PLP	CU	Total	Child	Parent	Difficulty	Comfort
CLP	-							
PLP	.40*	-						
CU	.41*	.57**	-					
Total	.70***	.80***	.84***	-				
Child	.52**	.22	.46*	.52**	-			
Parent	.37	.76***	.78***	.83***	.38	-		
Difficulty	.32	.59**	.66***	.73***	.26	.75***	-	
Comfort	.60**	.48*	.38	.58**	-.01	.16	.09	-

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

Table 7

*TORQ Planned Comparisons between Observation Settings*

Scale	Mean (SD)		<i>F</i>	Cohen's <i>d</i>
	All (N=27)	Clinic (n=15) Home (n=12)		
CLP Total	6.22 (1.85)	6.47 (1.96) 5.92 (1.73)	0.58	0.31
PLP Total	8.00 (2.47)	8.80 (2.57) 7.00 (2.00)	3.96	0.80
CU Total	7.37 (2.66)	7.60 (2.59) 7.08 (2.84)	0.24	0.20
Total React.	27.67 (7.20)	28.93 (6.96) 26.08 (7.46)	1.05	0.41
Child Rep.	5.85 (2.11)	6.00 (1.89) 5.67 (2.42)	0.16	0.16
Parent Rep.	8.07 (3.06)	8.53 (2.53) 7.50 (3.66)	0.75	0.35
Difficulty	6.44 (2.10)	6.80 (2.37) 6.00 (1.71)	0.97	0.40
Comfort	7.26 (3.47)	7.53 (2.92) 6.92 (4.17)	0.04	0.18

All planned comparisons  $p > 0.05$ .

Table 8

*DPICS Interrater Reliability by Observation Setting*

Code	Intraclass Correlation			Fisher r-to-Z	<i>p</i> (two-tailed)
	Full Sample (N=27)	Clinic (n=15)	Home (n=12)		
<i>Child Codes</i>					
NTA	.96	.96	.87	1.43	.15
CM	.91	.89	.80	0.76	.45
QU	.99	.99	.95	1.90	.06
PRO	.96	.94	.96	-0.49	.62
YE	.84	.90	.82	0.74	.46
WH	.94	.98	.88	2.15	.03*
CO	.88	.83	.94	-1.28	.20
NC	.95	.71	.98	-3.29	.00**
NOC	.94	.86	.98	-2.35	.02*
AN	.93	.94	.92	0.35	.73
NA	.67	.67	.68	-0.04	.97
NOA	.85	.73	.93	-1.70	.09
<i>Parent Codes</i>					
NTA	.94	.90	.96	-1.04	.30
DC	.95	.96	.95	0.25	.80
IC	.92	.92	.92	0	1
LP	.89	.68	.92	-1.66	.10
UP	.96	.97	.95	0.57	.57
DQ	.98	.96	.99	-1.54	.12
IQ	.98	.94	.99	-1.99	.05

RF	.77	.71	.84	-0.73	.47
BD	.68	.64	--	--	--
TA	.91	.83	.92	-0.88	.38

---

\* $p < 0.05$ ; \*\* $p < 0.01$ ; reliability scores based on a stratified random sampling of 30 DPICS segments such that each DPICS segment (CLP, PLP, and CU) and observation setting (clinic, home) are equally represented. -- = ICC is incalculable because this code did not occur during the segments selected for reliability analysis.

Table 9

*DPICS Means and Standard Deviations during 5-Minute CLP by Observation Setting*

Code	Mean (SD)		<i>F</i>	<i>d</i>
	Total (N=27)	Clinic (n=15) Home (n=12)		
<i>Child Codes</i>				
NTA	0.81 (1.24)	0.93 (1.28) 0.67 (1.23)	0.30	0.21
CM	4.56 (4.26)	5.20 (3.80) 3.75 (4.83)	0.76	0.35
QU	5.26 (5.53)	6.40 (4.05) 3.83 (6.89)	1.46	0.49
PRO	30.52 (13.32)	34.87 (13.67) 25.08 (11.12)	4.01	0.81
CO	2.33 (2.80)	2.00 (1.65) 2.75 (3.84)	-0.47	-0.28
NC	0.22 (0.58)	0.13 (0.35) 0.33 (0.78)	-0.79	-0.36
NOC	2.70 (3.20)	2.40 (2.20) 3.08 (4.21)	-0.30	-0.22
AN	3.56 (2.78)	3.60 (2.41) 3.50 (3.29)	0.01	0.04
NA	0.78 (1.05)	0.53 (0.92) 1.08 (1.17)	-1.89	-0.55
NOA	1.33 (1.41)	0.87 (0.92) 1.92 (1.73)	-4.12	-1.06
YE	0.15 (0.46)	0.07 (0.26) 0.25 (0.62)	-1.08	-0.41
WH	0.07 (0.39)	0.13 (0.52) 0 (0)	0.79	0.35
Inappro	1.15 (1.63)	1.13 (1.77) 1.17 (1.53)	-0.00	-0.02
CORatio <sup>A</sup>	0.89 (0.26)	0.90 (0.28) 0.88 (0.23)	0.04	0.10
NCRatio <sup>A</sup>	0.10 (0.26)	0.10 (0.28) 0.09 (0.22)	0.02	0.06
<i>Parent Codes</i>				
NTA	0.70 (1.07)	0.80 (1.27) 0.58 (0.79)	0.27	0.21
DC	3.59 (5.81)	2.80 (1.78) 4.58 (8.59)	-0.62	-0.32
IC	1.74 (1.85)	1.73 (1.83) 1.75 (1.96)	-0.00	-0.01

LP	0.33 (0.62)	0.40 (0.63)	0.25 (0.62)	0.38	0.25
UP	2.37 (1.62)	2.53 (1.96)	2.17 (1.12)	0.33	0.23
IQ	5.67 (3.83)	5.00 (2.83)	6.50 (4.82)	-1.02	-0.41
DQ	10.63 (5.36)	11.27 (5.42)	9.83 (5.41)	0.47	0.28
RF	1.85 (1.49)	2.00 (1.46)	1.67 (1.56)	0.33	0.23
BD	0.15 (0.36)	0.13 (0.35)	0.17 (0.39)	-0.05	-0.11
TA	28.04 (12.02)	27.27 (9.82)	29.00 (14.73)	-0.13	-0.15
DO Skills	4.70 (2.57)	5.07 (2.99)	4.25 (1.96)	0.67	0.33
DON'T Skills	17.04 (7.09)	17.07 (7.15)	17.00 (7.34)	0.00	0.01

All  $p > .05$ ; <sup>A</sup>n = 22 (Clinic n=13, Home n=9); compliance and noncompliance ratios were incalculable due to the parent not giving compliable commands during the observation segment (i.e., the parent gave no commands or gave commands with which the child could not comply)

Table 10

*DPICS Means and Standard Deviations during 5-Minute PLP by Observation Setting*

Code	Mean (SD)			<i>F</i>	<i>d</i>
	Full Sample (N=27)	Clinic (n=15)	Home (n=12)		
<i>Child Codes</i>					
NTA	1.85 (3.05)	2.07 (3.24)	1.58 (2.91)	0.16	0.16
CM	2.48 (1.93)	3.20 (1.78)	1.58 (1.78)	5.49*	0.95
QU	5.37 (5.61)	5.47 (3.96)	5.25 (7.38)	0.01	0.04
PRO	31.07 (12.68)	33.27 (14.61)	28.33 (9.67)	1.01	0.41
CO	6.48 (4.54)	6.40 (3.78)	6.58 (5.52)	-0.01	-0.04
NC	1.63 (2.60)	1.27 (1.53)	2.08 (3.55)	-0.65	-0.32
NOC	9.15 (6.55)	8.73 (6.20)	9.67 (7.20)	-0.13	-0.15
AN	3.70 (3.24)	3.87 (2.72)	3.50 (3.92)	0.08	0.12
NA	0.85 (1.23)	0.53 (0.74)	1.25 (1.60)	-2.38	-0.62
NOA	2.07 (2.13)	2.47 (2.36)	1.58 (1.78)	1.15	0.44
YE	0.67 (1.82)	0.27 (0.46)	1.17 (2.66)	-1.68	-0.52
WH	0.33 (1.21)	0.20 (0.56)	0.50 (1.73)	-0.40	-0.25
Inappro	2.85 (5.19)	2.53 (3.16)	3.25 (7.11)	-0.12	-0.14
CORatio	0.81 (0.20)	0.81 (0.22)	0.80 (0.19)	0.01	0.04
NCRatio	0.18 (0.18)	0.17 (0.18)	0.20 (0.19)	-0.19	-0.17
<i>Parent Codes</i>					
NTA	1.74 (2.67)	1.93 (3.35)	1.50 (1.57)	0.17	0.16
DC	10.33 (9.25)	8.73 (6.26)	12.33 (12.01)	-1.01	-0.40
IC	6.96 (5.10)	7.73 (6.03)	6.00 (3.67)	0.76	0.35



LP	0.41 (1.01)	0.27 (0.46)	0.58 (1.44)	-0.65	-0.32
UP	2.70 (2.66)	3.20 (3.12)	2.08 (1.88)	1.19	0.44
IQ	6.63 (5.26)	6.87 (4.96)	6.33 (5.82)	0.07	0.10
DQ	9.07 (5.82)	10.07 (6.19)	7.83 (5.31)	0.98	0.40
RF	2.04 (1.81)	2.07 (1.58)	2.00 (2.13)	0.01	0.04
BD	0.15 (0.60)	0.27 (0.80)	0 (0)	1.33	0.47
TA	31.74 (9.49)	32.40 (7.49)	30.92 (11.84)	0.16	0.16
DO Skills	5.44 (3.23)	5.80 (3.36)	5.00 (3.13)	0.40	0.25
DON'T Skills	17.44 (9.98)	18.87 (9.78)	15.67 (10.37)	0.68	0.33

---

\* $p < 0.05$

Table 11

*DPICS Means and Standard Deviations during 5-Minute CU by Observation Setting*

Code	Mean (SD)			<i>F</i>	<i>d</i>
	Full Sample (N=27)	Clinic (n=15)	Home (n=12)		
<i>Child Codes</i>					
NTA	1.78 (3.58)	2.60 (4.63)	0.75 (0.97)	1.84	0.55
CM	1.33 (1.62)	2.00 (1.85)	0.50 (0.67)	7.08*	1.07
QU	5.52 (5.37)	8.00 (5.83)	2.42 (2.47)	9.57**	1.24
PRO	16.74 (11.65)	18.20 (8.74)	14.92 (14.74)	0.52	0.29
CO	7.70 (4.61)	7.47 (4.87)	8.00 (4.45)	-0.09	-0.12
NC	3.15 (3.93)	2.27 (2.22)	4.25 (5.28)	-1.75	-0.53
NOC	10.89 (10.53)	9.93 (10.40)	12.08 (11.04)	-0.27	-0.21
AN	0.63 (0.93)	1.00 (1.07)	0.17 (0.39)	6.55*	1.03
NA	0.52 (0.94)	0.47 (1.13)	0.58 (0.67)	-0.10	-0.12
NOA	0.74 (0.90)	0.73 (1.10)	0.75 (0.62)	-0.00	-0.02
YE	0.74 (1.93)	0.40 (1.12)	1.17 (2.62)	-1.05	-0.41
WH	1.70 (2.52)	1.73 (2.74)	1.67 (2.35)	0.00	0.02
Inappro	4.11 (6.80)	4.53 (8.06)	3.58 (5.09)	0.13	0.14
CORatio <sup>A</sup>	0.76 (0.22)	0.78 (0.19)	0.73 (0.26)	0.24	0.20
NCRatio <sup>A</sup>	0.24 (0.24)	0.22 (0.19)	0.27 (0.26)	-0.24	-0.20
<i>Parent Codes</i>					
NTA	2.37 (2.31)	2.67 (2.26)	2.00 (2.41)	0.55	0.30
DC	12.85 (10.15)	10.20 (6.77)	16.17 (12.79)	-2.43	-0.63
IC	8.89 (6.58)	9.47 (6.55)	8.17 (6.83)	0.25	0.20

LP	0.19 (0.48)	0.20 (0.41)	0.17 (0.58)	0.03	0.06
UP	3.85 (3.57)	3.93 (3.79)	3.75 (3.44)	0.02	0.05
IQ	1.89 (1.72)	2.20 (2.11)	1.50 (1.00)	1.11	0.42
DQ	4.70 (3.84)	5.47 (4.21)	3.75 (3.25)	1.35	0.47
RF	0.70 (1.03)	0.73 (1.16)	0.67 (0.89)	0.03	0.06
BD	0.33 (0.83)	0.60 (1.06)	0 (0)	3.85	0.79
TA	26.30 (13.75)	30.87 (12.87)	20.58 (13.10)	4.19	0.82
DO Skills	5.07 (4.60)	5.47 (5.11)	4.58 (4.03)	0.24	0.20
DON'T Skills	8.96 (5.52)	10.33 (6.57)	7.25 (3.39)	2.17	0.59

\* $p < 0.05$ ; \*\* $p < 0.01$  ; <sup>A</sup>n = 29 (Clinic n=14, Home n=12); compliance and noncompliance ratios were incalculable due to the parent not giving compliable commands during the observation segment (i.e., the parent gave no commands or gave commands with which the child could not comply)