

Rowan University

## Rowan Digital Works

---

Theses and Dissertations

---

7-19-2012

### The examination of inter-rater reliability and the reliability of attachment measures

Nicole Bayles

Follow this and additional works at: <https://rdw.rowan.edu/etd>



Part of the [Child Psychology Commons](#), and the [Student Counseling and Personnel Services Commons](#)

Let us know how access to this document benefits you - share your thoughts on our [feedback form](#).

---

#### Recommended Citation

Bayles, Nicole, "The examination of inter-rater reliability and the reliability of attachment measures" (2012). *Theses and Dissertations*. 16.  
<https://rdw.rowan.edu/etd/16>

This Thesis is brought to you for free and open access by Rowan Digital Works. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Rowan Digital Works. For more information, please contact [LibraryTheses@rowan.edu](mailto:LibraryTheses@rowan.edu).

**THE EXAMINATION OF INTER-RATER RELIABILITY AND THE  
RELIABILITY OF ATTACHMENT MEASURES**

by  
Nicole M. Bayles

A Thesis

Submitted to the  
Department of Psychology  
College of Liberal Arts and Sciences  
In partial fulfillment of the requirement  
For the degree of  
Master's of Arts in School Psychology  
at  
Rowan University  
April 12, 2012

Thesis Chair: Dr. Roberta Dihoff, Ph.D.



## **Dedications**

*I would like to dedicate this thesis to my mother, Susan Bayles, without whom none of this  
would have been possible!*

## **Acknowledgement**

I would like to thank Dr. Roberta Dihoff and Dr. Terri Allen for all of their help and guidance throughout my work on this project.

## **Abstract**

Nicole Bayles

### **THE EXAMINATION OF INTER-RATER RELIABILITY AND THE RELIABILITY OF ATTACHMENT MEASURES**

2011/12

Roberta Dihoff, Ph.D.

Master of Arts in School Psychology

The purpose of this study was to determine the reliability of measuring attachment through subjective measures, as well as the inter-rater reliability of those using the subjective measures. Participants were volunteers who had worked with the Division of Youth and Family Services (DYFS). They included social workers, lawyers, psychologists, etc. Participants were asked to answer questionnaires regarding the DYFS children they had worked with, as well as demographics about themselves. The questionnaires also included an alternative informed consent. The surveys were mailed to participants and included return envelopes to ensure ease of mailing them back. It was hypothesized that measuring attachment subjectively is not a reliable form of measure and that there is no significant inter-rater reliability. Results showed no reliability of subjective measurements of attachment and no inter-rater reliability, supporting the hypothesis. The study will be a stepping stone toward a standard of measure and reliability when assessing attachment.

## Table of Contents

Abstract.....	v
List of Figures.....	vii
Chapter 1: Introduction.....	1
Chapter 2: Literature Review.....	3
Chapter 3: Methodology.....	20
Chapter 4: Results.....	22
Chapter 5: Discussion.....	25
List of References.....	28
Appendix A: Questionnaire.....	32

## List of Figures

Figure	Page
Awareness of Standard Measures.....	22
Use of Attachment Measures.....	23
Use of Different Measures Per Child.....	24
Psychologist Agreement of Attachment Per Child.....	24



## **Chapter 1**

### **Introduction**

#### **Need for the Study**

Attachment theory was developed by John Bowlby and Mary Salter Ainsworth between 1950 and 1960 (Bretherton, 1992). To date, researchers have used attachment theory in numerous studies to help explain later relationship issues or mental health conditions. Attachment theory has also been used by New Jersey's Division of Youth and Family Services (DYFS) as a way to determine an unhealthy household and possible removal of a child. Due to the subjective nature of the theory, finding a reliable way to measure it has become an issue. Also, there has not been much research conducted on the reliability of the assessments or inter-rater reliability in these subjective tests. Now, the question becomes whether a subjective approach is reliable enough or if a more standardized measure should be used to determine attachment.

#### **Purpose and Hypotheses**

The purpose of the study was to determine the reliability of assessing attachment through subjective measures as well as inter-rater reliability. It was hypothesized that measuring attachment subjectively is not a reliable form of measure and that there is no significant inter-rater reliability.

#### **Assumptions**

This study assumed that the DYFS social workers were qualified and knowledgeable about the children to accurately answer the survey about attachment and measurement.

## **Limitations**

Although there were great successes in the study, limitations were sometimes inevitable. For this study, the limitations included the convenience sample obtained through DYFS. The convenience sample was a limitation within itself, but it also brought other limitations. The convenience sample was small and the findings may not be generalized. Also, the convenience sample may only have provided for information on one demographic and the study may have different results in a different sample. Lastly, the use of self-report data has limited validity.

## **Summary**

Attachment theory originated in the 1950s (Bretherton, 1992) and lacks a standardized measure. The Division of Youth and Family Services relies on psychologists' subjective measures to determine a child's attachment and possible removal from their parents. The current study will measure the reliability of those subjective measures as well as inter-rater reliability. The researcher hypothesizes that measuring attachment subjectively is not a reliable form of measure and that there is no significant inter-rater reliability. This study is a necessary step towards finding a reliable measure of assessing attachment. The hope is that removal of a child will be based on a reliable measure of attachment with no question in the minds of the social workers or psychologists.

## **Chapter 2**

### **Literature Review**

#### **Attachment Theory Overview**

Attachment theory was developed to describe the bond that forms between infants and their parents or primary caregiver (Bowlby, 1958). Bowlby believed that infants have instinctual drives, e.g. sucking and crying, that form the core of the attachment to their caregiver (Bowlby, 1958). The reaction of the caregiver to these drives gives the infant expectations about future interactions. The outcomes of those future interactions will form the type of attachment for the infant (Bowlby, 1958). While some researchers believe that the form of attachment will remain constant unless changed by therapy or some type of trauma, others believe that as people grow and change so may their attachment formation (Baldwin & Fehr, 1995; Fox, 1995; Lewis, Feiring & Rosenthal, 2000; Main, Kaplan, & Cassidy, 1985; Weinfield, Sroufe & Egeland, 2000). The attachment of the infant has been found to be strongest between 8 months and 3 years of age and most obvious when the child is in a frightening situation (Bowlby, 1958; 1982).

Since the beginning of attachment theory, the research on infant and parent attachment has continued, and research on adolescent and adult attachment has begun (Ciechanowski, Walker, Katon, & Russo, 2002; Difilipo & Overholser, 2002). However, all of the research on the many types of attachment through the lifespan has brought many different measures for assessing attachment at different stages. Due to this, three different psychologists could potentially use three different measures for assessing the attachment of a child. The different types of measures of attachment include observations (Ainsworth, 1978; Crittenden, 1992), interviews (George, Kaplan & Main, 1985),

questionnaires (Bartholomew & Horowitz, 1991; Hazan & Shaver, 1987), Q-sort tests (Vaughn & Water, 1990; Waters & Deane, 1985) and Stem stories (Emde, Wolf & Oppenheim, 2003; Green, Stanley, Smith, & Goldwyn, 2000).

Each type of measure of attachment claims to find different forms of attachment and results in different types of scores, such as continuous scores or discrete categories (George & West, 1999; Werner-Wilson & Davenport, 2003). The type of attachment a child may form will depend on the measure used. There are some measures, like the Strange Situation measure, that will result in three different types of attachment, which are secure, avoidant, and anxious/ambivalent (Ainsworth, 1978; Hazan & Shaver, 1987). Other measures of attachment, such as the Attachment Interview for Childhood and Adolescents (AICA), result in four types of attachment categorized as secure/autonomous, dismissing, preoccupied, and uninvolved/disorganized (Main, Kaplan & Cassidy, 1985). Lastly, there are other measures of attachment that believe attachment should result in a continuous scale rather than a fixed form like the categories mentioned above (Waters & Dean, 1985).

There is also the issue of inter-rater reliability between the different assessments. In a study by Hadadian, Tomlin and Sherwood-Puzzello (2005), mental health professionals, who are trained to practice developmental assessment and intervention and testify in custody cases, were surveyed to determine their level of comfort in practicing such assessment and intervention and their interest in continuing their training. The survey measured these aspects in areas of attachment, behavior, and regulation/adaptation (Hadadian et al., 2005). The results of the study found that the mental health professionals felt comfortable in their knowledge of attachment, but were significantly

more interested in behavior and adaptation. Also, the professionals believed that behavior and adaptation were more important for families' common needs. Professionals also stated that behavior and adaptation training gave professionals more concrete ways of helping parents through play, discipline, sleep, and feeding (Hadadian et al., 2005). The results also showed that only non-Caucasian professionals found significant interest in attachment and believed families could benefit from it (Hadadian et al., 2005). This study shows that the inter-rater reliability in attachment assessment may be compromised by using different professionals to assess one child (Hadadian et al., 2005).

In assessing such results, the conclusion can be made that Caucasian professionals may be biased towards different measures, such as behavior, and may not rate a child's attachment the same way as a non-Caucasian professional. For a custody battle, one professional may rate the child as being more securely attached to the father and another professional may find the opposite. The issue then becomes deciding who is correct and where the child should be placed.

## **Types of Attachment Assessment**

### **Observations**

As previously stated, attachment measures can be broken down into different categories. Originally, Mary Ainsworth developed an assessment in which an infant is simply observed during different situations with varying amounts of stress (Ainsworth, 1978). This is called the Strange Situation assessment of attachment. During this form of assessment, a child is allowed time to play. A researcher observes the child as their caregiver, as well as other strangers, leave and then re-enter the room (Ainsworth, 1978). Ainsworth believed that this was a way to recreate the daily interactions a child may

experience with their caregivers and other strangers. This form of assessment observes the amount of exploration by the child and the child's reactions when their caregiver leaves and returns (Ainsworth, 1978). The reliability of this assessment can be found through many different studies since its development (Antonucci & Levitt, 1984; Main, Kaplan, & Cassidy, 1985; Wartner et al., 1994). However, this form of assessment is not appropriate for children over the age of 18 months, which called for new assessments needed for older toddlers and older children.

The need for a measure of older children brought about the creation of the Preschool Assessment of Attachment (PAA) by Crittenden (1992). The PAA was designed for children 18 months to five years old, when the Strange Situation is no longer applicable. Like the Strange Situation, the PAA uses coded observation. However, the PAA was modified to fit patterns that children develop after 18 months of age. These patterns are titled Type A – defended, Type B – secure or balanced, and Type C – coercive. These patterns refer to the ways in which children, at these ages, approach their interpersonal relationships (Crittenden, 1992). This measure is also designed to distinguish between normal and obsessive, compulsive behavior patterns that could lead to emotional or behavioral issues in the future (Crittenden, 1992).

The validity of the PAA was tested by Teti and Gelfand (1997) and, more recently, by Crittenden, Claussen, & Kozłowska (2007). Also, in a study by Fairchild (2006), the validity and reliability of many different assessments were measured. The PAA was one assessment in this study. Fairchild (2006) found that the reliability of the PAA ranged from 80-90% inter-rater agreement. Fairchild (2006) also report that the PAA showed content and concurrent validity.

## **Q-Sort Tests**

A second form of measuring attachment is called a Q-sort test. Waters and Dean (1985) developed this form of methodology for observation of children ages one to five years old. This methodology allows for observation in varying environments, instead of only in a lab. The Q-sort test provides observers with 100 descriptive statements designed to cover the entire spectrum of attachment related behaviors (Waters, 1995). These behaviors include the secure base behaviors, exploratory behavior, social cognition, and affective responses (Waters & Dean, 1985). The Q-sort test allows psychologists to place a child on a continuum between securely attached and insecurely attached. However, the Q-sort test does not allow for subcategories of insecure attachment (Waters & Dean, 1985). Today, psychologists are using the Attachment Q-set Version 3.0 which was last updated in 1987 (Waters, 1995). The reliability and validity of this type of measure can be found in many studies (Bretherton et al., 1990; Howes & Ritchie, 1999; Teti et al., 1996; Vaughn & Waters, 1990). In a study by Fairchild (2006), reliability was found to range from 72-97% inter-rater agreement. The study also found the Q-sort assessment to have both content and concurrent validity (Fairchild, 2006).

## **Interviews**

The third type of attachment measure is the attachment interview. The first type of attachment interview developed was the Adult Attachment Interview by George, Kaplan and Main (1985). This form of measure focused on adolescent and adult attachment. The interview prompts participants to reflect on their childhood and how those experiences may have impacted their own personality or behavior in adolescents or adulthood. The answers to the interview are coded and a style of attachment is determined (George et al.,

1985). The reliability and validity of this type of assessment can be found in a study completed by Bakermans-Kranenburg & van IJzendoorn (1993). This study found the reliability of the assessment using a test- retest method. Through this study, the researchers found the reliability to be 78% and also found strong discriminant validity (Bakermans-Kranenburg & van IJzendoorn, 1993).

The AAI was later reorganized for middle childhood. The Child Attachment Interview (CAI) was designed by Target, Fonagy, and Schmueli-Goetz (2003). This form of interview was developed for children between the ages of seven and eleven years old. Adapted for children from the Adult Attachment Interview (AAI), this measure focuses on the representations of a child's relationships with their caregivers and events related to attachment. The researcher uses both verbal and nonverbal communication to score the interview. The CAI results are reported in four different styles of attachment (Target et al., 2003).

Five years after its development, the researchers came together again to test the reliability and validity of Child Attachment Interview (Schmueli-Goetz, Target, Fonagy, & Datta, 2008). The study used a test-retest design to measure reliability and validity (Schmueli-Goetz et al., 2008). The researchers found the test-retest results to remain stable between the two tests with the reliability ranging between 69% and 85% (Schmueli-Goetz et al., 2008). They also found the discriminant validity to be strong with no outside variables skewing the classification results (Schmueli-Goetz et al., 2008). The researchers also measured inter-rater reliability and found that after only three days of training, interviewers could reliably code the assessment (Schmueli-Goetz et al., 2008).



A third interview type of assessment is the Attachment Interview for Childhood and Adolescence (AICA). This is another form of assessment developed from the AAI for children between the ages of eight and twelve. This study was developed by Ammaniti, van IJzendoorn, Speranza, & Tambelli (2000). Like the Adult Attachment Interview, this measure assesses a child's state of mind regarding attachment instead of relationship quality. The questions are coded to result in a style of assessment for the child (Ammanti et al., 2000).

During the development of this assessment, the researchers measured its reliability and validity. They used a test-retest method at ages 10 and 14 years. They found the results to be stable with a reliability ranging from 50-78% between tests (Ammanti et al., 2000). Also, the construct validity is strong since the results do not differ from results of similar assessments such as the Adult Attachment Interview, from which the AICA was adapted (Ammanti et al., 2000). There is also strong inter-rater reliability since it is so close to the AAI and anyone trained in AAI can reliably code this assessment also (Ammanti et al., 2000).

### **Questionnaires**

The fourth form of measurement for attachment is questionnaires. The first questionnaire was developed by Hazan and Shaver (1987). This questionnaire was developed for measuring attachment in adults. The questionnaire places adults into the three main categories formed by Ainsworth in her original development of attachment (Ainsworth, 1978). The questionnaire provides statements for the participants and they are asked to choose which one best describes how they feel. The average of the

statements chosen would result in one of three styles of attachment for that participant (Hazan & Shaver, 1987).

During the development of this assessment, Hazan and Shaver (1987) conducted five separate studies to determine which forms of questions best represented a person's attachment style. Through the study they found that questions asking about early relationships with parents were best predictive of attachment style (Hazan & Shaver, 1987). However, the researchers did discuss that self-report questionnaires are not always the best form of assessment and finding reliability and validity is difficult (Hazan & Shaver, 1987). Yet, the development of new questionnaires continued in pursuit of making attachment assessment easier.

The second questionnaire for attachment is called the Relationship Questionnaire (RQ-CV) and was developed by Bartholomew and Horowitz (1991). This questionnaire was based on the idea that attachment styles are reflections of a person's thoughts about their partner and themselves. The questionnaire is based on a person's positive or negative feelings in both areas. The combinations of those feelings about self and others result in placing the participant in one of four attachment styles (Bartholomew & Horowitz, 1991). The researchers also replicated the original study to find reliability and validity. They found construct validity by comparing the results of the study to the results of others and found that they did not significantly differ (Bartholomew & Horowitz, 1991). They also found inter-rater reliability throughout the coding of the questionnaires (Bartholomew & Horowitz, 1991).

The last forms of questionnaires are the Experiences in Close Relationships (ECR) and the Experiences in Close Relationships – Revised (ECR-R). The original

questionnaire was developed by Brennan, Clark and Shaver (1998). The revision of this questionnaire was later developed by Fraley, Waller and Brennan (2000). These questionnaires group items into two dimensions of attachment, anxiety and avoidance (Fraley et al., 2000). These two categories place participants into four attachment styles by combining the amount of avoidance and anxiety one has based on the questions (Fraley, et al., 2000). For example, someone who has low anxiety and low avoidance would be rated as securely attached (Fraley et al., 2000).

In 2005, the reliability and validity of the Experiences in Close Relationships – Revised (ECR-R) assessment was studied (Sibley, Fischer, & Liu, 2005). This study ran 3 separate test- retest methods. The results of the studies showed strong convergent and discriminant validity (Sibley et al., 2005). They also found stability and reliability (85%) from test to test (Sibley et al., 2005).

### **Stem Stories**

The final type of attachment measure is called Stem stories. This measure of attachment can be used for children anywhere between three and eight years of age. It uses dolls and narrative to enact a story. In the beginning of the measure, the interviewer uses the dolls to begin the story and then hands them to the child to complete the story. The interviewer may use varied degrees of encouragement or prompting to help the child if needed. These stem stories are designed to find the child's inner working models of attachment relationships. One of these types of measures is called the Manchester Child Attachment Story Task (MCAST) which was developed by Green, Stanley, Smith, and Goldwyn (2000). A study by Fairchild (2006) found that the MCAST showed a range of

80-94% inter-rater agreement and reliability. The MCAST also has strong content and discriminant validity (Fairchild, 2006).

All of these different types of measures, from observations to questionnaires, make it difficult to decide which measure is best. The decision becomes even more difficult when the reliability and validity of some assessments is hard to find or nonexistent at all. If a Division of Youth and Family Services (DYFS) worker called three different psychologists to determine the removal of a child, each psychologist could use a different measure, with different reliability or validity, and have a different conclusion. This makes attachment a risky determinant of removal. Due to this, there have been many court cases in which attachment has been argued for and against.

#### **Attachment Theory within Custody Cases**

In many custody cases, the goal is to find a solution for the child's stress reduction and healthy attachment to at least one parental figure (George, Isaacs, & Marvin, 2011). Also in many cases, it can become easy to lose sight of the child and their attachment needs. Keeping this in mind is most important for children under five who are most vulnerable to separation and the outcome of custody decisions (George et al., 2011). Because of this, determining a child's attachment to both parents may be used as an important decision maker for custody decisions.

In the Family Court Review for July, 2011, George et al. (2011) discusses a case of a 2 year old and her parents' custody case. Even in a case in which attachment is used to help the case, the authors point out that there is not a uniform training or single assessment that mental health professionals use to measure attachment (George et al.,

2011) The authors also point out that there are many professionals who are certified to measure attachment after only a two and half hour class (George et al., 2011).

Although there is only minor research completed on attachment in divorce or custody, it is still used in such cases. The authors of this article believe that any research on attachment can be generalized and used in custody and divorced cases (George et al., 2011). For example, the authors state that the populations of parents who are separated and going through a divorce share the same problems as other high risk populations that have been researched (George et al., 2011). The authors also discuss the strong internal validity of attachment measures, but do not discuss reliability of such measures (George et al., 2011). Also, in custody disputes, there is no standard for visitation schedules, so assessing attachment must be tailored to each family (George et al., 2011). This may cause some unreliability in assessment because duration and frequency of assessment will be different for every child and may be different for the father and mother of the same child. For example, if a mother is given full custody of a child and the father only granted visitation every other weekend, the child will be spending a lot more time with the mother and, therefore, attachment measures may not fairly judge attachment to the father and may favor the mother incorrectly.

For this case, the two year old child's attachment was measured through both the Strange Situation assessment and a Stem story assessment using dolls (George et al., 2011). The child's parents were also assessed for attachment through the Adult Attachment Interview. The results of the attachment measures found that the child's attachment to her father bordered between secure and ambivalent. The child's attachment to her mother bordered between secure and avoidant (George et al., 2011). In the end, the

professionals used these results to recommend joint custody for the child. Since the child was borderline with both parents, there was no necessity to grant full custody one way or the other.

In a second article presented in the Family Court review, Main, Hesse, & Hesse (2011) discuss some of the aspects that must be taken into account when using attachment in the court systems. First, they discuss that the assessments should be mandated by the court (Main et al., 2011). In previous cases, each parent would present their own “expert” witness which adds to more unreliability of the results presented by professionals. If the court was to mandate and over see such assessments, the reliability of the results would increase (Main et al., 2011). Also, the court should choose the individuals to administer the assessments to assure the professionals are highly trained and reliable so inter-rater reliability is not an issue (Main et al., 2011). Second, the researchers/ authors of this article push for a scientific base for testimony as opposed to the “expert testimony” which was usually an opinion. Also, the researchers take a position shared with other researchers that many of the assessments used in custody cases have problems which make the quality and reliability of their results questionable (Main et al., 2011; O’Donohue, Beitz, & Tolle, 2009). Third, the researchers discuss the three attachment assessments which are considered to be the “gold standard measures.” This means that these tests have extremely high validity and reliability, and the findings of each test have been established in multiple laboratories and studies (Main et al., 2011). These three measures are the Strange Situation, the Adult Attachment Interview, and the Attachment Q-Sort. The researchers/ authors of this article believe that courts should only mandate the use of one of these assessments dependent on the age of the individual that

must be assessed (Main et al., 2011). This could allow the use of attachment to become more reliable in removal of a child and in the courtroom.

A third Family Court Review article by Schmidt, Cuttress, Lang, Lewandowski and Rawana (2007) discusses the use of attachment theory in cases of maltreated children. The authors discuss the difficulty in assessing such situations because abusive parents cannot accurately discuss their parent-child relationship or risky behaviors (Schmidt et al., 2007). Also, the authors discuss the downfalls of attachment assessment. As discussed previously, attachment assessment is very subjective. There can be too few observations used, or a lack of agreement on how to assess parent-child relationships. More importantly, there is no agreement on what is the most important or relevant aspects in the relationship (Schmidt et al., 2007). When using attachment in abusive cases of maltreatment in children, the authors believe that it is important to not only assess the child but also the parent. Since assessing attachment in children usually happens during times of child distress, it is important to also examine the parent's ability to respond to the child in such times of distress and need (Schmidt et al., 2007). They also discuss the importance of multiple, home assessments. The authors believe that home observations are important in order to allow the parents to act comfortably without instruction. However, there is not even a standard for home observation assessments. Also, the authors suggest longer visits due to the fact that brief observations do not place enough demand on the parents to interact with the child or be emotionally available for the child (Schmidt et al., 2007). Lastly, because caretaking activities are most critical in the formation of secure attachments, observations should be made during times such as feeding, sleeping, or bathing (Schmidt et al., 2007). All of these things need to be taken

into consideration when assessing a child's attachment for use in court. It is important to provide the most accurate and reliable results, especially if there is the possibility of removing a child from their parent(s).

Through all of these court documents, one general pattern emerges. The use of attachment in the court system is tricky and care must be taken to properly use attachment in the courtroom. In the Family Court Review, Jennifer McIntosh interviewed Everett Waters, the co-founder of the New York Attachment Consortium and an expert on Bowlby and Ainsworth, to find out what questions need to be considered and answered when using attachment in disputes about child custody (Waters & McIntosh, 2011). Waters and McIntosh (2011) believe that attachment is highly misunderstood and overused. Waters believes that people should just say what they mean instead of just jumping to the word attachment. He states that, when speaking about attachment, it is important to make sure people are being specific enough about what they are saying and that they are not over-utilizing the word itself (Waters & McIntosh, 2011). Waters goes on to explain that it is also important to remove misconceptions about attachment, such as the "strength of attachment." Waters says that there is no such concept as a strong or weak attachment and that assessments do not measure strengths, only the type of attachment (Waters and McIntosh, 2011). Also, the common misconception about a window of opportunity for forming attachment is incorrect. So what does this mean for custody? Waters states that what the courts may deem as the right place for a child now may not be the best solution as time progresses (Waters and McIntosh, 2011).

Another important discussion with this article is attachment assessment within custody cases. Waters tell McIntosh that the use of attachment assessment should not be



primary. He believes that having “convergent evidence” is important, but he would not jump to attachment as the first measure for a child in a custody dispute (Waters and McIntosh, 2011). Waters states that sometimes the insecure forms of attachment that children develop can be very detrimental, however it is not always clear what causes the problems or how it may be fixed through custody arrangements (Waters and McIntosh, 2011). Waters states, “I do not think that disorganized attachment in an infant, or unresolved attachment in an adult, in and of itself, would be the sole basis for a court decision” (Waters and McIntosh, 2011).

Another concept important to decision of removal of a child based on attachment is the child’s attachment to other non-nuclear family figures. An article by Riggs (2003) discusses the drawbacks of only focusing on attachment to parents when assessing children in the court systems. Riggs states that society and the court systems typically believe that the nuclear family is most important, however, this is not the standard family dynamic. There are many children, today, living in single parent homes, remarried families, foster homes, homosexual homes, and homes composed of multiple generations living in one household (Riggs, 2003). Also, such alternative families are becoming the norm, while the traditional nuclear family is falling out of favor (Riggs, 2003). Additionally, current society causes the need for a multiple income household, forcing parents to find non-parent caregivers for the children (Riggs, 2003). Many children today are cared for by a grandparent, aunt, neighbor, close friend, or stranger in a childcare setting (Riggs, 2003). Due to this shift in care giving from parents to multiple people, it is only fair to assume that a child will form multiple attachments encompassing multiple forms of attachment which will effect their development (Riggs, 2003). Due to this, it is

important for government agencies or court systems to assess attachment in unconventional relationships and not only parent-child relationships.

### **Conclusion**

To conclude, attachment can be seen as a highly misunderstood topic that should not be taken lightly, whether that be in the court systems or in government agencies. The extensive amount of ways in which a psychologist could decide to assess attachment further obscures the reliability of the theory's use in court room decisions. Currently, New Jersey's Division of Youth and Family Services (DYFS) uses attachment theory and assessment in the decision to remove a child from a home. If a child needs to be assessed for attachment, psychologists are asked to choose their own assessment and report back to the DYFS case manager for that child. The issue lies in the fact that there is no standard for the measure used and, therefore, the results can be different for the same child. This calls for the need of a standardized measure within DYFS.

Also, the literature on the use of attachment focuses on custody decision in divorce cases, some of this literature may be generalized for use in removal of a child as the same concepts for attachment apply. For example, court systems are calling for a standard of measurement when assessing attachment in a child (Main, Hesse & Hesse, 2011). Also, attachment should be measured outside of the nuclear family as this is not the standard family in current society (Riggs, 2003). However, the lack of research on reliability of the use of attachment in removal of a child, as well as the lack of inter-rater reliability in such cases does call for the necessity of this current study. It is possible that the drawbacks and unreliability of the use of attachment in court decisions of custody can

be generalized to removal of child, but it is also probable that there are major differences which must be taken into consideration and a new study must be completed.

## **Chapter 3**

### **Methodology**

#### **Participants**

The participants in this study were 6 individuals who have completed or arranged for multiple bonding evaluations in New Jersey and volunteered to participate. They consisted of 5 females and 1 male, ranging in age from 20 to 61 years old. The participants were from multiple ethnic backgrounds with a majority being Caucasian (66.7%). They also ranged in years they have been involved with bonding evaluations, from 1 to 10 years and 21 to 30 years.

The children described in the survey, by the participants, had a large range of ages. The youngest children described were under 2 years of age and the oldest were over 14 years of age. Also, the children came from a wide range of ethnicities. Both male and female children were described by the participants.

#### **Materials**

The materials in this study consisted of a survey, which included an informed consent statement, as well as a demographic questionnaire. (See Appendix A)

#### **Design**

The current study was a descriptive study. Only descriptive information and statistics were provided.

#### **Procedure**

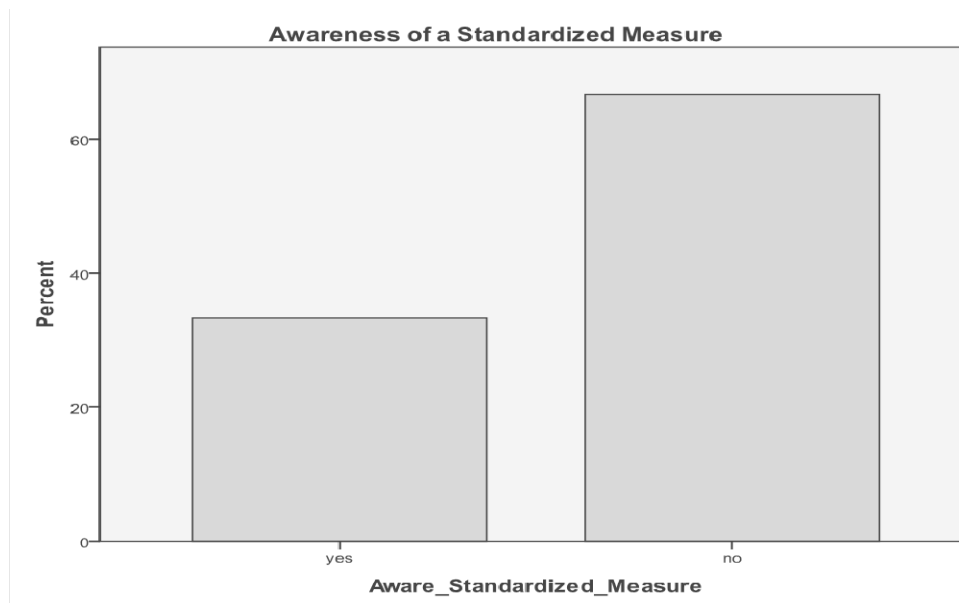
A survey and demographics questionnaire was mailed out to 10 participants for completion. Sixty percent returned the survey. After the survey and demographic questionnaire had both been filled out, the researcher asked that the participants mail the

surveys back using the return envelope provided to ensure anonymous results. The researcher then placed the surveys into a secured folder and placed them in the locked cabinet until it was time for analysis.

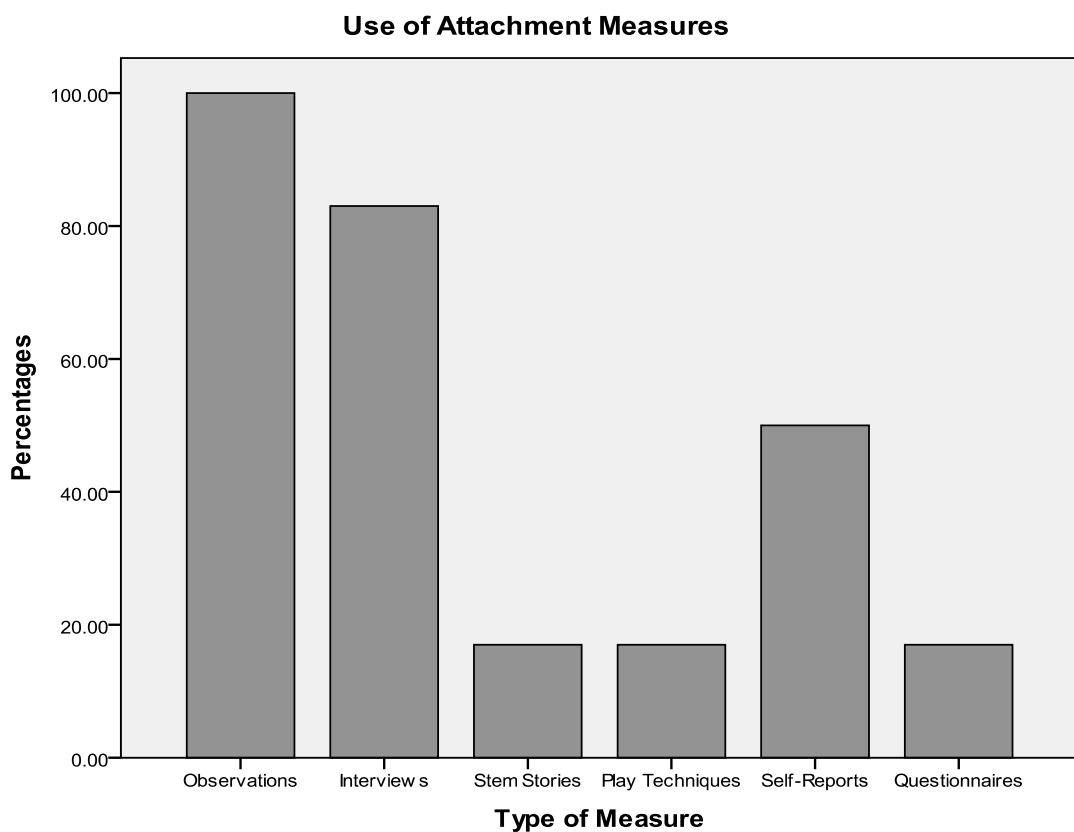
## Chapter 4

### Results

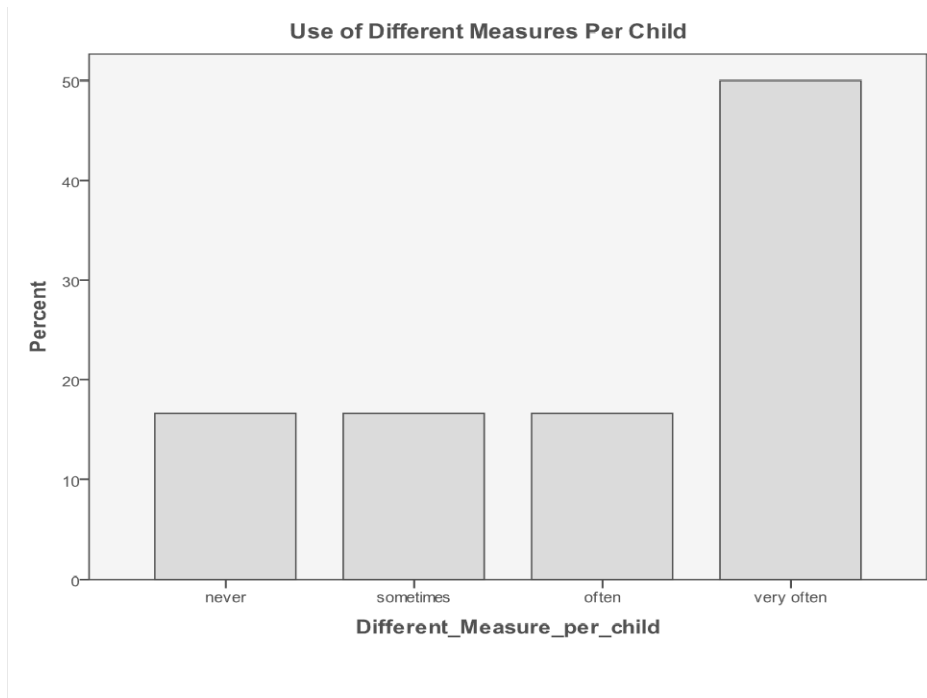
It was hypothesized that measuring attachment subjectively is not a reliable form of measure and that there is no significant inter-rater reliability. The results showed that two out of six participants (33.3%) has seen and were aware of any kind of “standardized instrument.” These participants assumed that a bonding evaluation was a “standardized instrument,” however, it is not. This assumption alone can cause issues.



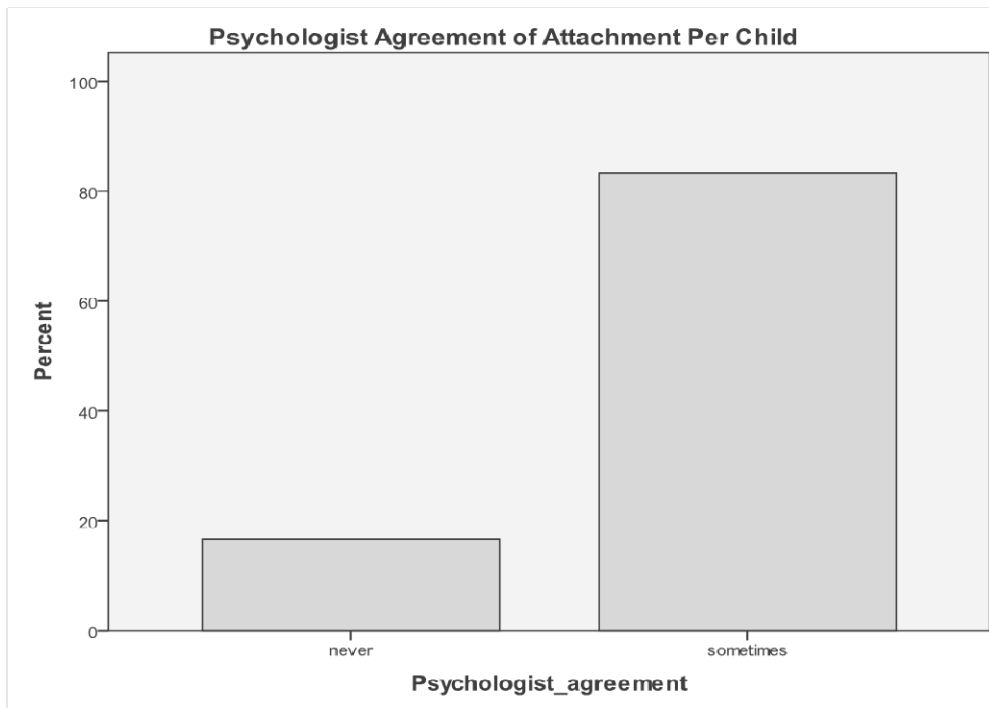
Participants in this study also described a range of types of measures used to assess the children. There are six types of assessments that may be used to measure attachment. Participants reported that observation is always used, followed by interviews (83%) and self-report measures (50%). Yet, the other three types of assessment are rarely used.



The study's results also showed that different measures are very often used on each child (50%). Yet, none are standard measures. This can cause issues because each measure can have a different conclusion. However, a positive outcome of this study is that, usually, there is more than one family member assessed as well. It is not just the biological parents, but also foster parents, step parents, and legal guardians.



Lastly, results showed that psychologists only agree on the attachment of the child sometimes (83%) or never (17%). There were no participants that answered often or always.





## **Chapter 5**

### **Discussion**

#### **Summary**

Attachment theory was developed in the 1950's to describe the bond that forms between infants and their parents or primary caregiver (Bowlby, 1958). The attachment of the infant has been found to be strongest between 8 months and 3 years of age and most obvious when the child is in a frightening situation (Bowlby, 1958; 1982). Over time, research has expanded in order to examine attachment in older children, adolescents, and adults (Ciechanowski, Walker, Katon, & Russo, 2002; Difilipo & Overholser, 2002). This research is where all of the different types of attachment measures come from. However, there has never been any kind of standardization for use of these measures in custody situations.

Even in a court case in which attachment is used to help the case, the authors point out that there is not a uniform training or single assessment that mental health professionals use to measure attachment (George et al., 2011) The authors also point out that there are many professionals who are certified to measure attachment after only a two and half hour class (George et al., 2011). Main, Hesse, and Hesse (2011) discuss that the assessments should be mandated by the court (Main et al., 2011). If the court was to mandate and over see such assessments, the reliability of the results would increase (Main et al., 2011).

The current study sought to find support for the beliefs of the current literature. The purpose of the study was to determine the reliability of assessing attachment through subjective measures as well as inter-rater reliability. It was hypothesized that measuring

attachment subjectively is not a reliable form of measure and that there is no significant inter-rater reliability. The results showed that there is no standard for assessment. With no standard, it is not possible to have inter-rater reliability, which was also shown by the results. All of the results within this study do support the hypothesis and support the current literature.

### **Implications**

The findings of this study imply that a move towards a standard when measuring attachment is necessary. It is also important to not place so much weight on subjective results. By advancing the literature, this study shows how important it may be to form a standard for assessing attachment when using the results in custody battles or removal of a child from their home by an agency such as DYFS. With even more research completed in this area, it could mean a lot for psychologists working with the DYFS system. Instead of choosing their own form of assessment, they would have to follow a standard. This would be better for the children and their families because conclusions between psychologists would be less likely to not agree, which this study showed was happening.

### **Limitations**

Although this study resulted in important findings about attachment assessment, limitations were sometimes inevitable. For this study, the limitations were found in the convenience sample. The convenience sample was a limitation within itself, but it also brought other limitations. The convenience sample was small and the findings may not be able to be generalized. Also, the convenience sample may only have provided for information on one demographic and the study may have different results in a different sample, such as a family service agency in Kentucky. Lastly, the use of self-report data

has limited validity and confusion about questions can occur resulting in skewed data. Also, this study was purely descriptive.

### **Future Directions**

This study shows a great need for further research. Future studies should try to use a larger and more diverse pool of participants. With more participants and more data, it could give even more insight into the issues surrounding the use of attachment assessment without a standard. The next step after that would be to have variables and experimental data. Experimenting with a standard assessment versus any assessment could show a significant difference in the reliability of assessment results, as well as inter-rater reliability. Using experimental data would also alleviate the limited validity of self-report data.

## References

- Ainsworth, M. D. (1967). *Infancy in Uganda: Infant Care and the Growth of Love*. Baltimore, MD: Johns Hopkins Press.
- Ainsworth, M. D. (1978) *Patterns of Attachment: A Psychological Study of the Strange Situation*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Antonucci, T. & Levitt, M. (1984). Early prediction of attachment security: A multivariate approach. *Infant Behavior and Development*, 7(1), 1-18.
- Ammanti, M., van IJzendoorn, M. H., Speranza, A. M., & Tambelli, R. (2000). Internal working models of attachment during late childhood and early adolescence: An exploration of stability and change. *Attachment and Human Development*, 2(3), 328-246.
- Baldwin, M. W., & Fehr, B. (1995). On the instability of attachment style ratings. *Personal Relationships*, 2(3), 247-261.
- Bakermans-Kranenburg, M., & van IJzendoorn, M. H. (1993). A psychometric study of the adult attachment interview: Reliability and discriminant validity. *Developmental Psychology*, 29(5), 870-879.
- Bartholomew, K., & Horowitz, L. M. (1991). Attachment styles among young adults: A test of a four-category model. *Journal of Personality & Social Psychology*, 61(2), 226-244.
- Bowlby, J. (1958). The nature of the child's tie to his mother. *The International Journal of Psychoanalysis*, 39, 350-373.
- Bowlby, J. (1969). *Attachment and Loss: Vol. 1. Attachment*. New York, NY: Basic Books.
- Bowlby, J. (1973). *Attachment and Loss: Vol. 2. Separation*. New York, NY: Basic Books.
- Bowlby, J. (1982). Attachment and loss: Retrospect and prospect. *American Journal of Orthopsychiatry*, 52, 664-678.
- Brennan, K. A., Clark, C.L., & Shaver, P. R. (1998). Self-report measurement of adult romantic attachment: An integrative overview. In J.A. Simpson & W.S. Rholes (Eds.), *Attachment Theory and Close Relationships* (46-76). New York, NY: The Guilford Press.
- Bretherton, I., Prentiss, C., & Ridgeway, D. (1990). Family relationships as represented in a story-completion task at thirty-seven and fifty-four months of age. *New Directions for Child and Adolescent Development*, 1990(48), 85-105

- Bretherton, I., & Munholland, K. A. (1999). Internal working models in attachment relationships: A construct revisited. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of Attachment: Theory, Research, and Clinical Applications* (89–111). New York, NY: The Guildford Press.
- Cassidy, J. (1986). The ability to negotiate the environment: An aspect of infant competence as related to quality of attachment. *Child Development*, 57, 331–337.
- Ciechanowski, P. S., Walker, E. A., Katon, W. J., & Russo, J. E. (2002). Attachment theory: A model for health care utilization and somatization. *Psychosomatic Medicine*, 64, 660-667.
- Crittenden, P. M. (1992). Quality of attachment in the preschool years. *Development and Psychopathology*, 4(2), 209–241.
- Crittenden, P. M., Claussen, A. H., & Kozłowska, K. (2007). Choosing a valid assessment of attachment for clinical use: A comparative study. *Australian and New Zealand Journal of Family Therapy*, 28(2), 78-87.
- Difilippo, J. M., & Overholser, J. C. (2002). Depression, adult attachment, and recollections of parental caring during childhood. *Journal of Nervous and Mental Disease*, 190, 663-669.
- Emde, R. N., Wolf, D. P., & Oppenheim, D. (2003). *Revealing the Inner Worlds of Young Children: The MacArthur Story Stem Battery and Parent-Child Narratives*. New York, NY: Oxford University Press, Inc.
- Fairchild, S. R. (2006). Understanding attachment: Reliability and validity of selected attachment measures for preschoolers and children. *Child and Adolescent Social Work Journal*, 23(2), 235-261.
- Fox, N. A. (1995). Of the way we were: Adult memories about attachment experiences and their role in determining infant-parent relationships: A commentary on van IJzendoorn (1995). *Psychological Bulletin*, 117(3), 404-410.
- Fraley, R.C., Waller, N.G., & Brennan, K.A. (2000). An item-response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, 78, 350-365.
- George, C., Isaacs, M. B., & Marvin, R. S. (2011). Incorporating attachment assessment into custody evaluations: The case of a 2-year-old and her parents. *Family Court Review*, 49(483). Madison, WI: Association of Family and Conciliation Courts.
- George, C., Kaplan, N., & Main, M. (1985). *Adult Attachment Interview*. Unpublished manuscript, University of California, Berkeley.
- George, C., & West, M. (1999). Developmental and social personality models of adult attachment and mental ill health. *British Journal of Medical Psychology*, 72, 285-303.

- Green, J., Stanley, C., Smith, V., & Goldwyn, R. (2000). A new method of evaluating attachment representations in young school-age children: The manchester child attachment story task. *Attachment and Human Development*, 2(1), 48-70.
- Hadadian, A., Tomlin, A. M., & Sherwood-Puzzello, C. M. (2005). Early intervention service providers: What do they say about their infant mental health training needs? *Early Child Development and Care*, 175, 431-444.
- Hazan, C., & Shaver, P. (1987). Romantic love conceptualized as an attachment process. *Journal of Personality and Social Psychology*, 52(3), 511-524.
- Howes, C., & Richie, S. (1999). Attachment organizations in children with difficult life circumstances. *Development and Psychopathology*, 11, 251-268.
- Lewis, M., Feiring, C., & Rosenthal, S. (2000). Attachment over time. *Child Development*, 71, 707-720.
- Main, M., Hesse, E., & Hesse, S. (2011). Attachment theory and research: Overview with suggested applications to child custody. *Family Court Review*, 49(426). Madison, WI: Association of Family and Conciliation Courts.
- Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood, and adulthood: A move to the level of representation. *Monographs of the Society for Research in Child Development*, 50(1-2), 66-104.
- Main, M., & Solomon, J. (1990). Procedures for identifying infants as disorganized/disoriented during the Ainsworth strange situation. *Attachment in the Preschool Years: Theory, Research and Intervention* (121-160). Chicago, IL: The University of Chicago Press.
- McIntosh, J. (Interviewer) & Waters, E. (Interviewee). (2011). Are we asking the right questions about attachment [Interview Transcript]. Retrieved from *Family Court Review*, 49(474). Madison, WI: Association of Family and Conciliation Courts.
- O'Donohue, W., Beitz, K., & Tolle, L. (2009). Controversities in child custody evaluations. In J. L. Skeem, K. S. Douglas, & S. O. Lilienfeld (Eds.), *Psychological science in the courtroom: Consensus and controversy*. New York, NY: The Guilford Press.
- Riggs, S.A. (2003). Response to Troxel v. Granville: Implications of attachment theory for judicial decisions regarding custody and third-party visitation. *Family Court Review*, 41(39). Thousand Oaks, CA: Sage Publications, Inc.
- Sibley, C. G., Fischer, R., Liu, J. H. (2005). Reliability and validity of the revised experiences in close relationships (ECR-R) self-report measure of adult romantic attachment. *Personality and Social Psychology Bulletin*, 31(11), 1524-1536.

- Sroufe, L. A., & Fleeson, J. (1986). Attachment and the construction of relationships. In W. Hartrup & Z. Rubin (Eds.), *Relationships and Development* (51–71). Hillsdale, NJ: Erlbaum.
- Sroufe, L. A., & Waters, E. (1977). Attachment as an organizational construct. *Child Development*, 48, 1184–1199.
- Target, M., Fonagy, P., & Schmueli-Goetz, Y. (2003) Attachment representations in school-age children: The development of the Child Attachment Interview (CAI). *Journal of Child Psychotherapy*, 29(2), 171-186.
- Teti, D. M., & Gelfand, D. M. (1997). The Preschool Assessment of Attachment: Construct validity in a sample of depressed and nondepressed families. *Development and Psychopathology*, 9, 517-536.
- Teti, D. M., Sakin, J. W., Kucera, E., Corns, K. M., & Das Eiden, R. (1996). And baby makes four: Predictors of attachment security among preschool-age firstborns during the transition to siblinghood. *Child Development*, 67, 579–596.
- Vaughn, B. E., & Waters, E. (1990). Attachment behavior at home and in the laboratory: Q-sort observations and strange situation classifications of one-year-olds. *Child Development*, 61(6), 1965-1973.
- Wartner, U. G., Grossmann, K., Fremmer-Bombik, E., & Suess, G. (1994). Attachment patterns at age six in south Germany: Predictability from infancy and implications for preschool behavior. *Child Development*, 65, 1014-1027.
- Waters, E. (1995). Appendix A: The attachment q-set (Version 3.0). *Monographs of the Society for Research in Child Development*, 60(2-3), 234-46.
- Waters, E., & Deane, K. E. (1985). Defining and Assessing Individual Differences in Attachment Relationships: Q-Methodology and the Organization of Behavior in Infancy and Early Childhood. *Monographs of the Society for Research in Child Development*, 50(1-2), 41-65.
- Weinfield, N. S., Sroufe, L., & Egeland, B. (2000). Attachment from infancy to early adulthood in a high-risk sample: Continuity, discontinuity, and their correlates. *Child Development*, 71(3), 695-702.
- Werner-Wilson, R. J., & Davenport, B. R. (2003). Distinguishing between conceptualization of attachment: Clinical implication in marriage and family therapy. *Contemporary Family Therapy*, 25, 179-193.

## Appendix A

### Questionnaire

The purpose of this survey is to evaluate the reliability of assessment measures and the inter-rater reliability of those administering the measures. The research, entitled "Examination of Inter-rater Reliability and the Reliability of Attachment Measures" is being conducted by Nicole Bayles of the Psychology Department, Rowan University, in fulfillment of her M.A. degree in School Psychology. For this study you will be required to answer survey questions regarding the children you are or have been responsible for as a DYFS worker. Your participation in the study should not exceed 20 minutes. There are no physical or psychological risks involved in this study, and you are free to withdraw your participation at any time without penalty.

The data collected in this study will be analyzed and will be submitted for approval of graduation. Your responses will be anonymous and all the data gathered will be kept confidential.

By taking this survey you agree that any information obtained from this study may be used in any way thought best for publication or education provided that you are in no way identified and your name is not used.

Participation does not imply employment with the state of New Jersey, Rowan University, the principal investigator, or any other project facilitator.

If you have any questions or problems concerning your participation in this study, please contact Nicole Bayles at (908) 278-7798, Bayles70@students.rowan.edu, or her faculty advisor, Dr. Roberta Dihoff, [Dihoff@rowan.edu](mailto:Dihoff@rowan.edu).

*Please answer the following questions to the best of your ability.*

1. How many children under your supervision have been assessed for attachment?

- |          |                  |
|----------|------------------|
| a. None  | d. 50-100        |
| b. 1-20  | e. More than 100 |
| c. 20-50 |                  |

2. Please indicate the amount of children assessed per age group below.

- |             |               |
|-------------|---------------|
| a. 0-2 yrs: | d. 6-8 yrs:   |
| b. 2-4 yrs: | e. 8-10 yrs:  |
| c. 4-6 yrs: | f. 10-12 yrs: |



g. 12-14 yrs:

h. 14+ yrs:

3. Please indicate the amount of children assessed per ethnicity below.

a. Caucasian:

f. Arab/ Arab American:

b. Black/ African American:

g. Multiracial:

c. Asian/ Asian American:

h. Other: (please list)

d. Hispanic/ Latino:

e. American Indian or

Eskimo:

4. Please indicate the amount of each gender of all of the children.

\_\_\_\_\_ Males

\_\_\_\_\_ Females

5. Are you aware of any standardized instruments used for assessing attachment?

a. Yes

i. If yes, please list: \_\_\_\_\_

b. No

6. What kind of measure did the psychologists use for each child? (Please indicate

how many children per type of measure)

a. Observation:

d. Play Techniques:

b. Interview:

e. Self-Report

c. Stem Stories

f. Questionnaires:

7. Please list how many psychologists usually assess each child?

- a. 1
- b. 2
- c. 3
- d. 4
- e. 5+

8. How often do the psychologists each use a different measure on the same child?

(1=Never, 3=Sometimes, 5=Very Often)

1      2      3      4      5

9. How often do the psychologists agree on the attachment of each child? (1=Never,

3=Sometimes, 5=Very Often)

1      2      3      4      5

10. How long is each child typically evaluated by each psychologist? (Ranges may be used)

11. Other than the child, who else is usually evaluated?

- a. Biological mother
- b. Biological father
- c. Step parents
- d. Other relatives
- e. Foster parent
- f. Guardian
- g. Other (Please list)

12. Have you ever seen bonding evaluations:

a. Yes

i. If yes, approximately how many times: \_\_\_\_\_

b. No

***Please provide your opinion for the following items***

1. Do you believe the evaluation of each child's behavior was: (Please indicate number of children for each)
  - a. Too Short:
  - b. Normal:
  - c. Too Long:
2. Do you believe the behaviors observed were: (Please indicate number of children for each)
  - a. Normal for the child:
  - b. Only situational:
  - c. Both:
3. Do you believe the use of different measures per child may cause unreliable results?
  - a. Yes
  - b. No
  - c. Sometimes
  - d. Unsure
4. Do you believe the different cultures of children may skew results:
  - a. Yes
  - b. No
  - c. Sometimes
  - d. Unsure

## Demographic Questionnaire

*Please circle the response that best fits you.*

1. Please indicate your age range.

- a. 20-30
- b. 31-40
- c. 41-50
- d. 51-60
- e. 61 or older

2. Please indicate your ethnicity.

- a. Caucasian/ White
- b. Black/ African American
- c. Asian/ Asian American
- d. Hispanic/ Latino
- e. American Indian or Eskimo
- f. Arab/ Arab American
- g. Multiracial
- h. Other (Please list):

3. Gender:        Male        Female

4. Number of years working with the DYFS system.

- a. 1-10 years
- b. 11-20 years
- c. 21-30 years
- d. More than 30 years