

Running head: PREDICTORS OF QUALITY IN FORENSIC EVALUATIONS

AN EXPLORATION OF PREDICTORS OF QUALITY IN FORENSIC PSYCHOLOGICAL
EVALUATIONS IN THE FIELD OF CHILD WELFARE

A DISSERTATION SUBMITTED TO THE FACULTY
OF
THE GRADUATE SCHOOL OF APPLIED AND PROFESSIONAL PSYCHOLOGY
OF
RUTGERS,
THE STATE UNIVERSITY OF NEW JERSEY
BY
ERICA T. LEE
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE
OF
DOCTOR OF PSYCHOLOGY

NEW BRUNSWICK, NEW JERSEY

AUGUST 2020

APPROVED:

N. Andrew Peterson, Ph.D.

Nancy Harrison, Psy.D.

DEAN:

Francine Conway, Ph.D.

PREDICTORS OF QUALITY IN FORENSIC EVALUATIONS

Copyright © 2020 by Erica Lee

ABSTRACT

Despite the growing knowledge base about best practices in forensic evaluations in general, there is a current lack of literature focused on high-quality forensic evaluations in the field of child welfare. As child welfare evaluations have an impact on particularly vulnerable populations, it is paramount for measures of quality within these evaluations to be defined, examined, and measured. The purpose of this study was to contribute to the developing literature regarding forensic evaluations in child welfare by describing factors that lead to high-quality psychological evaluations and high-quality recommendations. This study examined a sample of 895 forensic evaluations in child welfare as a secondary data analysis of data that were originally collected by the New Jersey Coordination Center for Child Abuse and Neglect (NJCC). The concept of quality in this dissertation was defined as an evaluation that adheres to the objectives of New Jersey's Department of Children and Families' (DCF) Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (2012). The variables in this study were measured by the Quality Improvement (QI) Tool, an instrument designed by the NJCC that supports forensic evaluation training, peer review, and supervision. A series of logistic regression models were analyzed with the following predictor variables: the presence of a diagnostic assessment, the integration of culture, and the clarity of the referral question. Subject age category and gender were also included in the analyses. It was found that the presence of a diagnostic assessment, integration of culture, the clarity of the referral question, and subject age category were all significant predictors of overall quality. Additionally, it was found that variable interaction terms were not significant predictors of overall quality. Results also showed that the presence of a diagnostic assessment and subject age category were significant predictors of quality of recommendations.

Again, it was found that the variable interaction terms were not significant predictors of quality of recommendations. Implications for clinical practice are discussed, along with recommendations for future research.

ACKNOWLEDGMENTS

I am indebted to my committee members, Dr. N. Andrew Peterson and Dr. Nancy Harrison, who have assisted me in setting goals and provided me with a wealth of knowledge and support that has inspired me to obtain these goals. I would like to acknowledge Dr. Peterson's expertise around best research practices and willingness to support my project in the first place, as well as Dr. Harrison's commitment to advances in the field of child welfare and unwavering positivity. This would not have been possible without the special skills and knowledge that each of you graciously bestowed upon me.

Thank you to the New Jersey Coordination Center of Child Abuse and Neglect (NJCC) for allowing me to use this data in order to conduct research to further to field. The work conducted at the NJCC is meaningful and impactful, and I am so glad I was able to be a part of it. Thank you especially to Matthew Forsythe for always being willing to help me with various parts of the project and for encouraging me to learn more about statistical methods. Thanks to Christina Diaz, my fellow research assistant, for always being there for support as we both went through the writing process.

Finally, I would like to acknowledge my partner Dan, as well as my family and friends. I would have never been able to get through my doctoral studies without Dan's unwavering support. I am also eternally grateful for all of the support and understanding from family and friends throughout this process.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	ix
INTRODUCTION	1
Overview of the History of Child Maltreatment	2
Explanatory Models: The Etiology of Child Maltreatment	4
Sexual abuse	5
Child neglect	5
Child physical abuse	8
Flawed theories of child physical abuse	9
Risk Factors for Child Maltreatment	10
The Psychologist's Role in Child Abuse and Neglect Evaluations	11
The Gap in the Current Literature	14
Quality Assurance and Improvement in Forensic Practice	17
The Current Study	21
Research Questions	22
METHOD	23
Data collection	23
Procedures	25
Sample	25
Measures	26

QI Tool: Code Book.....	27
Peer-review process	27
Reliability	28
Outcome Variables.....	29
Overall quality	29
Quality of recommendations	30
Predictors	33
Presence of a diagnostic assessment.....	33
Integration of culture	36
Clarity of referral question.	38
Analytic Approach	40
Handling missing values	41
RESULTS.....	42
Descriptive statistics	42
Correlations	43
Crosstabulations.....	44
Multinomial logistic regression analysis.....	47
Relationship between the presence of a diagnostic assessment and overall quality	51
Relationship between integration of culture and overall quality.....	52
Relationship between the clarity of the referral questions and overall quality.....	52
Relationship between age category and overall quality	53

Binomial logistic regression analysis.....	53
Relationship between the presence of a diagnostic assessment and quality of recommendations.....	54
Relationship between age category and quality of recommendations.....	55
DISCUSSION.....	55
Relationship between the presence of a diagnostic assessment and overall quality	57
Relationship between integration of culture and overall quality.....	59
Relationship between the clarity of the referral questions and overall quality.....	61
Relationship between covariates and overall quality.....	63
Relationship between the presence of a diagnostic assessment and quality of recommendations.....	65
Relationship between covariates and quality of recommendations	66
Specificity of recommendations.....	66
Study Impact on Forensic Psychological Evaluations.....	68
Portuguese forensic psychological reports.....	68
Forensic mental health evaluations on youth in the Netherlands.....	71
Quality of forensic reports submitted to Hawaii judiciary	73
Limitations of the Study	76

Strengths of the Study	77
Clinical Implications	78
Ethical Implications.....	81
Implications for Future Research	82
REFERENCES	85
APPENDICES	97
DCF's Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings.....	97
Quality Improvement Tool	114

LIST OF TABLES

Table	Page #
1. Descriptive Analysis of Frequency Counts and Percentages.....	43
2. Correlations.....	44
3. Crosstabulations for Overall Quality.....	45
4. Crosstabulations for Quality of Recommendations Recoded	47
5. Multinomial Logistic Regression Results with Odds Ratios	49
6. Predictors' Unique Contributions in the Multinomial Logistic Regression ..	51
7. Binomial Logistic Regression Results with Odds Ratios	54
8. Predictors' Unique Contributions in the Binomial Logistic Regression	54

Introduction

In the United States there are approximately four million reports of child abuse and neglect each year. As of 2016, State child protective services (CPS) agencies reported a referral rate of alleged child maltreatment of 55.1 referrals per 1000 children in the population (Child Welfare Information Gateway, 2018). Statistics on child maltreatment are voluntarily provided to the National Child Abuse and Neglect Data System (NCANDS) for the annual Child Maltreatment report by each state in the United States. Even though individual state's screening policies and responses to child maltreatment may differ, they are all based on the same standards set by federal law.

The Centers for Disease Control and Prevention (CDC) defines child maltreatment as any act, intentional or not, that results in harm, the potential for harm, or the threat of harm to a child (CDC, 2014). There are four broad types of child maltreatment measured by NCANDS; neglect, physical abuse, sexual abuse, and "other" which includes psychological abuse, threatened abuse/neglect, parental drug/alcohol addiction, and lack of supervision. Currently, approximately three quarters (74.8%) of reported alleged abuse in 2016 was for neglect, with 18.2% of cases being physical abuse, 8.5% of cases being sexual abuse, and 6.9% of cases being classified as "other." Many allegations included more than one form of alleged abuse. These reported cases of alleged abuse were typically submitted by professionals (64.9%) including education personnel, legal and law enforcement personnel, and social services personnel. Nonprofessionals comprised 18.1% of reporters of abuse, and this category includes friends, neighbors, and relatives. The remaining percentage of reporters wished to remain anonymous or were classified as "other."

Overview of the History of Child Maltreatment

Child maltreatment has technically occurred for a long time but has not always been described as such. It is only relatively recently that this topic has emerged as an area of rigorous study by researchers and policy makers. Typically, child- abuse- related terminology is considered to have emerged in the middle of the past century. However, there were some detailed descriptions of child abuse published in the French medical literature as early as the 19th century, and various other references throughout history. It is significant to understand how the topic of child abuse first generated scholarly interest, as abuse has always been happening, yet has not always been diagnosable or viewed as morally unacceptable.

Psycho-historical study of childhood in general reveals evidence that childhood in the past was routinely filled with terror, abuse, and neglect (deMause, 1988). It is speculated that this is because the concept of childhood did not exist until around the seventeenth century (Jackson, 1995). Instead, children were typically viewed as miniature adults and parenting was very detached. They were only given attention once it could be sure that they would live. Eventually, as infant mortality rates declined, children were seen as more important economically and as objects of affection (Raghavan & Alexandrova, 2015). This rise in affection and attention produced a new culture of childhood, and childhood finally became recognized as a stage of life (Ariès & van den Berg, 1978).

The development of the nuclear family and emerging recognition of childhood as a stage are crucial to understanding how child abuse may have originated. Children began attending schools, introducing a sharp divide between adult and child worlds, where the adult was seen as superior. Within the school setting, children were supervised more closely by their teachers,

which led to the evolution of corporal punishment as a popular means of discipline (Ariès & van den Berg, 1978).

The definition of child abuse is typically treated as obvious even though the actual composition of the meaning of the term is subject to enduring disagreement and differing opinions among people. Public perception of abuse and portrayal by the media seems to perpetuate the idea that there is a universal understanding of what constitutes child abuse. Philosopher Ian Hacking (1991) described humans as likely still doing the same vile things to children currently that they did a century ago. However, he explains that as people change their definitions of abuse they have been revising their values and moral codes as well. Because of these revisions, an action towards a child that was acceptable in the past to someone may currently be viewed by them as unacceptable or inhumane. Since the field of child maltreatment identification and intervention is rather new, there are still vast knowledge gaps and moral uncertainties regarding child abuse. People seem to change their definition of child abuse and maltreatment as their personal opinion changes, which is problematic (Mallon & Hess, 2014; Raghavan & Alexandrova, 2015). Research points out that as these definitions are vague, the underlying theories of what constitutes child well-being are vague as well.

As definitions of child maltreatment change, policy has progressed from an emphasis on child protection and child welfare to also incorporating child well-being, which involves intrapersonal, interpersonal, familial, and social well-being of children. Raghavan and Alexandrova (2015) attempted to define current theories of child well-being in the literature, as it is a concept that is constantly redefined by society. Their theory of child well-being is summarized as a child is doing well to the extent that they develop stage-appropriate capacities

that equip them for a successful future given their environment and that they engage with the world in child-appropriate ways, meaning with curiosity and emotional security.

It is important to examine how events in society have changed people's reactions to child maltreatment as well. Beginning in the 1960s, there was increased public attention on child welfare, when child abuse was finally defined as something morally wrong (Cradock, 2011). Moral, professional, and organizational interests combined at critical stages throughout history to publicize the issue of child abuse (Parton, 1979). With the invention of technology such as X-rays, child maltreatment became more detectable. The medical and legal fields initially struggled with the moral dilemma of whether or not it was more important to treat the child and family or punish the parents for hurting their children (Parton, 1979). Research indicates that even currently, social workers find it difficult to work within the field of child abuse and neglect and non-accidental injury, as there are many inconsistencies and therefore sometimes assumptions need to be made in regard to whether or not abuse occurred (Humphrey, 2015). This is often when forensic psychologists are called upon to perform evaluations on both alleged perpetrators and victims (Melton et al., 2017; Pence, 2011).

Explanatory Models: The Etiology of Child Maltreatment

Child maltreatment is an epidemic global health problem. As of 2011, the annual estimates of child maltreatment in the developed world were highest in the United States, where there were 3.4 million referrals made of alleged child abuse involving 6.2 million different children (Etter & Rickert, 2013). It seems intuitive that researchers should examine what is causing this reported child abuse and determine a way to prevent it. However, as of now, no specific etiologic factor of child abuse has been determined because causes and risk factors are

multifactorial. There are various models throughout the literature that attempt to explain why the different types of child abuse occur.

Sexual abuse. Finkelhor and Aranji (1986) attempt to explain the sexual abuse of children by adults using a four factor model. They conducted a review of the variety of theories that have been proposed to explain adults' sexual interest in children and noted that all of these theories appeared to be explaining one of four factors: emotional congruence, sexual arousal, blockage, or disinhibition. The first factor, emotional congruence, explains why a person would find relating sexually to a child to be emotionally gratifying and congruent. The next factor, sexual arousal, explains why an adult would be capable of being sexually aroused by a child. Blockage, the third factor, describes why a person would be frustrated or blocked in efforts to acquire sexual and emotional gratification from more normative sources. The final factor, disinhibition, attempts to explain why an adult would not be deterred by conventional social expectations and inhibitions against having sexual contact with a child. This four factor model illustrates that there are many theories of why sexual abuse of children occurs, and emphasizes that these factors apply on an individual and sociocultural level. Finkelhor and Aranji (1986) also posit that different types of inappropriate sexual behaviors towards children may need to be explained by different combinations of the four factors and the way that they interact.

Child neglect. Child neglect often receives less attention than other types of child maltreatment, even though it may be the most prevalent type of maltreatment (Dubowitz, 1999; Tzeng, Jackson & Karlson, 1991). In the literature this is referred to as the "neglect of neglect" (Garbarino & Collins, 1999). Neglect occurs on a continuum, often without clear points that aid in diagnosis, making it a challenge for clinicians, pediatricians, and social workers to address. Child neglect refers to the absence of behaviors, so it is difficult to measure and research as well

(Scannapieco & Connell-Carrick, 2002). It includes children not receiving medical or dental care, children who are inadequately fed or overfed, children who are not appropriately supervised, and children whose emotional and developmental needs are not sufficiently met (Dubowitz, 2014). Mothers who are accused of neglect are often socially disadvantaged in some way, and neglect often occurs in households headed by unmarried females (Sykes, 2011). The association between poverty and child maltreatment has been well-documented, but it is most evident among cases of child neglect (Sykes, 2011). One of the theories behind child neglect is that when mothers or parents in general are negatively labeled “neglectful” they will adopt strategies to minimize possible psychological consequences and insist on maintaining their perception of their identity as a good parent while distancing themselves from agencies that are likely to help them. This allows them to preserve their dignity but often can lead to hostile relationships with state agencies that require them to complete services (Sykes, 2011).

The eco-developmental theory regarding the etiology of child neglect and maltreatment states that it occurs due to the interaction of four different levels of systems and is influenced by the theories of Bronfenbrenner and Tinbergen. It is an ecological model that focuses on transactional and multi-level explanation of child maltreatment and neglect. These levels are called ontogenic, microsystem, exosystem, and macrosystem (Scannapieco & Connell-Carrick, 2002). The ontogenic level encompasses the childhood histories of abusive parents and attachment theories. Attachment theory posits that children who were maltreated have a possible predisposition to maltreat their own children. Research shows that the occurrence of a parent experiencing neglect in childhood alone is not enough to explain this parent in turn neglecting their own child (Hearn, 2011; Scannapieco & Connell-Carrick, 2002). However, this experience may influence future behavior and provide an additional context for maltreatment to occur

(Williams et al., 2011). Additionally, attachment theory posits that there is a reciprocal relationship between the parent and the child, and child characteristics such as temperament may affect this relationship. The caregiver may also be unresponsive to the child due to external factors such as domestic violence, substance abuse, psychological diagnoses, economic struggles, and dangerous living environments. These issues can occur at the microsystem level as well, which encompasses many of the additional factors that interact with the parents' developmental history. Within this level, child maltreatment and neglect is considered an interactive process (Scannapieco & Connell-Carrick, 2002; Williams et al., 2011).

Within the exosystem level there are several factors that have been correlated with child maltreatment and neglect within larger social structures such as work and neighborhood. For example, areas with more poverty and lower socioeconomic status have been correlated with child maltreatment. In neighborhoods with equal socioeconomic disadvantages, neighborhoods with social resources, whether formal or informal, child maltreatment occurred at a lower rate than within neighborhoods with fewer social resources (Tzeng, Jackson & Karlson, 1991). The exosystem has a lot of impact on the microsystem, as does the macrosystem. The macrosystem examines the embeddedness of the individual, family system, and community within the larger cultural setting (Scannapieco & Connell-Carrick, 2002). For example, the United States as a whole used to condone violence to some degree, and the line between physical abuse toward children and acceptable physical discipline was not clear (Hearn, 2011; Tzeng, Jackson, & Karlson, 1991).

Another theory regarding the etiology of neglect is the personalistic view of child neglect. This theory posits that parents' neglectful behavior is due to deficiencies in their personal characteristics. The apathy-futility syndrome represents a mixture of parental personality traits

such as low self-efficacy, inhibited emotional response, superficial relationships and loneliness, limited functional abilities, passive-aggressive expression of anger, reluctance or refusal to commit to positive beliefs, verbal inaccessibility to others, which limits problem solving to internal dialogues, and the ability to make those with whom they associate also feel futile (Tzeng, Jackson, & Karlson, 1991). Parents who are neglectful may also be impulsive, causing them to leave their offspring alone for extended periods of time to go fulfill their own needs. The personalistic view of child neglect places emphasis on the parents' difficulties that make them more susceptible to the stresses related to individual and social factors. An adequate parent may be able to handle these stressors, but neglectful parents cope by relying on disordered personality characteristics (Tzeng, Jackson, & Karlson, 1991).

Child physical abuse. There is an abundance of theories on the etiology of child physical abuse. Tzeng, Jackson, and Karlson (1991) describe 25 different theories in their book on the various theories of child abuse and neglect. Coohy and Braun (1997) attempted to construct an integrated conceptual framework for understanding child physical abuse based on a content analysis of these pre-existing theories. They identified three major determinants of physical abuse: parental exposure to aggression, parental exposure to stressors, and access to resources. At least two of three of these determinants are mentioned in each of the 25 theories cited above.

A review of theoretical approaches concludes that parental exposure to abuse in their own childhood may be an important factor for predicting child physical abuse. In particular, psychiatric and learning theories emphasize that parental exposure to aggression is a determinant of future perpetration of child abuse. Other theories also posit that members of the parent's social network influence the use of physical force within the family. Individual parental

exposure to stressors is hypothesized to be another determinant of child physical abuse (Coohey & Braun, 1997). Many theories state that the parent-child interaction may be a major source of stress that triggers physical abuse. Other theories focus on a broad range of stressors, both interpersonal and intrapersonal, that can affect parent behavior, such as divorce, separation from a loved one, interactions with employers, or the birth of a child. The last proposed determinant of physical abuse is access to resources, which can also be interpersonal or intrapersonal (Coohey & Braun, 1997). Many theoretical perspectives suggest that physically abusive parents lack both personal resources and psychological resources, such as knowledge of child development, social skills, and parenting skills. Because of these deficits, these individuals often use physical force to get their children to do what they want them to do (Tzeng, Jackson, & Karlson, 1991). Other types of resources these parents may be lacking include access to professionals outside their social circle such as social workers, teachers, child care workers, and lawyers, as well as access to economic resources such as from working or from illegal activities (Coohey & Braun, 1997).

Flawed theories of child physical abuse. Leventhal and Edwards (2017) describe inappropriate use of the theories behind child physical maltreatment by pediatricians, other medical personnel, attorneys, and journalists. They explain that there are three categories of flawed explanations of child abuse. The first category is when an abusive injury is explained as a legitimate injury, such as in an accidental fall or osteogenesis imperfecta, a rare illness. The next category includes other legitimate diagnoses that lack scientific support as explanations of injury, such as vitamin deficiencies or Ehlers-Danlos syndrome. The final category includes fabricated diagnoses that are purely speculative yet used to explain possible abuse, such as “dysphagic choking” and “temporary brittle bone disease.” Leventhal and Edwards (2017) urge

physicians, researchers, academic medical centers, journalists, and legal scholars to repudiate these scientifically unsupported theories that attempt to explain child physical abuse. They advise physicians who care for injured children to use a scientific approach and consider alternate hypotheses for their injuries while exercising careful clinical judgement in diagnosing abuse.

Risk Factors for Child Maltreatment

The total economic burden placed upon the United States due to cases of fatal and non-fatal child maltreatment is estimated to be around \$124 billion (CDC, 2014). In addition to this cost, many cases are not reported to social services and police and may cause a burden on children's future health. Child maltreatment has proven to have a negative effect on health problems in children and consequently these children are at a higher risk for health problems as adults. Some risk factors can be linked to an increase in the risk for the re-occurrence of abuse or neglect.

Although there is no specific etiologic factor of child maltreatment, both child and adult risk factors have been identified. Once risk factors are acknowledged, attempts can be made to prevent them from occurring and to increase protective factors. Parental risk factors include poor socioeconomic status, psychological problems, substance abuse, unmet emotional needs, and lack of parenting knowledge (Maguire-Jack & Font, 2017). Additionally, caregiver exposure to maltreatment as a child is a risk factor for abuse, as often abusive parents may be repeating the same type of child care practiced on them in their childhood (Kempe et al., 1985). It is not well understood why some caregivers, despite the presence of risk factors, are resilient and do not abuse children (CDC, 2014). Child risk factors include having physical or mental health problems, challenging behaviors, disabilities, excessive crying, frequent tantrums, low birth

weight, and being a twin (Maguire-Jack & Font, 2017). The way in which parents interact with their environment and community has implications for abuse and neglect, as higher levels of community involvement were found to be associated with lower levels of psychological aggression (Kim & Maguire-Jack, 2015). Additionally, children who are maltreated are disproportionately from low-income and racial minority families (U.S. Department of Health and Human Services, 2015).

As of 2016, after initial screening, approximately two thirds of the reported alleged incidents of child maltreatment in the United States were investigated (Child Welfare Information Gateway, 2018). Once these allegations, also known as referrals, are received, they are responded to in various ways. When appropriate, referrals receive responses from child welfare agencies that focus on addressing concrete need in the family, as the allegations do not warrant an investigation of abuse or neglect. However, a majority of these responses include an investigation that determines if a child was maltreated or is at-risk of maltreatment. This often includes a forensic psychological report involving the alleged perpetrator and/or the child. Intervention recommendations formulated by the psychologist are frequently included in this report (Child Welfare Information Gateway, 2018).

The Psychologist's Role in Child Abuse and Neglect Evaluations

The epidemic problems of children who are abused and neglected and their families create issues and questions that psychologists are called upon to address in a variety of ways. Often, psychologists are requested to complete forensic psychological evaluations to aid child welfare workers and the courts in making decisions regarding child protection matters. Psychological data and information gathered from these evaluations may provide a perspective

not otherwise available to the court, state agency, or other party involved, which makes these evaluations influential and significant (Pence, 2011).

Forensic psychology is the intersection between psychology and the law in that forensic psychologists can assist the legal system in decision making related to mental health and behavior (Zumbach, Wetzels, & Koglin, 2018). It should be acknowledged that forensic psychology is a subfield within clinical psychology. Sadoff and Dattilio (2012) documented that over 40 different subspecialties in the field of forensic mental health exist as well. Due to the nature of the relationship between the examiner and the examinee in the forensic context and how that can vary within different subspecialties, examiners must be aware of the various ethical implications in each unique situation. Forensic psychological evaluations differ from typical clinical psychological evaluations in many aspects, including the goals, scope, and product of the evaluation (Nicholson & Norwood, 2000).

As this study will be discussing forensic psychological evaluations specifically, it is important to distinguish the unique factors related to forensic evaluation procedures. Adults and children who receive forensic evaluations in the field of child welfare are referred for these evaluations by CPS. Sometimes the evaluation is court ordered, and other times it is requested of the parent or child in order to gain insight into recommendations that will be helpful to the family and CPS. In clinical evaluations, the client is typically the person who is being assessed, but in forensic evaluations, this is not the case (Mulay, Mivshek, Kaufman, & Waugh, 2018). In evaluations for child welfare purposes, the client is CPS because they are ordering and paying for the evaluations. In some cases, the person being evaluated may not want to participate but is forced to by the court, or feels obligated to because of the critical importance of the evaluation. This puts constraints on the limits of confidentiality (Melton et al., 2017). Any information

obtained in court ordered evaluations may be included in a report provided to the court. The person being evaluated in a forensic evaluation does not have much control over who reviews the evaluation report.

In clinical assessment settings, the psychologist is often in a helping role. Conversely, within the context of forensic psychology, psychologists are often seen as objective reporters (Greenberg & Shuman, 2007; Mulay et al., 2018). As the American Psychological Association's (APA) Guidelines for Psychological Evaluations in Child Protection Matters (2013a) state, psychologists are in a position to contribute significantly to decision-making in child protection. However, due to the conflicting desires mental health professionals might have to protect children from maltreatment while also helping to preserve families, forensic evaluations in the field of child welfare may be particularly challenging. Evaluators must remain neutral, unbiased, and impartial experts rather than becoming advocates and the tone of the evaluation is expected to reflect this (Young, 2016). Information provided from forensic evaluations should be intended to increase the fairness of decisions by the court, state agency, or other party (APA, 2013a).

Due to the critical importance of forensic evaluations, the subject is likely to alter their presentation in order to serve their interests. Numerous threats to validity exist in the forensic environment as compared to a clinical or therapeutic context. There is more of a potential for malingering, which is the exaggeration or feigning of mental health symptoms for an external gain, and faking good or bad, which is when someone acts in a certain way when they feel it will best serve their interests (Melton et al., 2017). Forensic evaluators must use tests that are sensitive to response style in order to aid in detecting malingering and faking good or bad (Young, 2016). The person being evaluated may also be reluctant to participate or tell the truth,

for fear of the consequences. Additionally, parents may feel the need to embellish or lie during interviews and assessments in order to try to protect themselves and/or their families. Melton et al. (2017) points out that other parties, such as lawyers, health care workers, police, and child welfare workers may distort the fact gathering processes involved in forensic evaluations by only presenting certain information to the examiner or presenting it in a biased manner. Forensic evaluators must examine information provided from as many perspectives as are available and must utilize data, assessment instruments, and behavioral observations to produce the most objective opinion possible (APA, 2013b). These evaluations must be focused and aim to answer referral questions, as opposed to clinical evaluations, which may be sweeping and broad (Allan & Grisso, 2014; Mulay et al., 2018).

APA's Guidelines for Psychological Evaluations in Child Protection Matters (2013a) detail the three stages characterizing the specific procedures that states use to guide their intervention in child protection cases. In the first phase, state CPS involvement is triggered by a report of suspected child maltreatment. In the following phase, if investigation results indicate that the child has been harmed or is at significant risk of being harmed, intervention strategies are typically applied. This can include voluntary services or protective child custody. In the final phase, the case may move from child protection to termination of parental rights and permanency planning for the child if efforts at reunification fail. The psychologist has a complicated role and can be asked to perform an evaluation during any of these phases.

The Gap in the Current Literature

Until the past fifteen years, forensic psychological evaluators had a limited amount of resources to use to improve their forensic report writing (Grisso, 2010). However, there have been several recent efforts to improve the quality of general forensic evaluations. Despite this

increase in the literature surrounding forensic evaluations, there is still a lack of information on child welfare evaluations in particular. It seems as if this has been a common problem, as back in 2000 Nicholson and Norwood mentioned a lack of studies scrutinizing the content and quality of child custody and parental capacity evaluations as well, which are highly related to child welfare. In Melton et al.'s (2017) book entitled *Psychological Evaluations for the Courts*, there is a chapter that discusses forensic report writing specifically in cases of child abuse and neglect. This is a unique resource that has many important recommendations that serve as a framework for improving forensic evaluations in general as well, including the suggestions to separate facts from inferences, stay within the scope of the referral question, avoid providing too much or too little information, and minimize clinical jargon. However, this resource is rare, and besides the specific guidelines published by the APA (2013a, 2013b) there is not an abundance of resources related to how to write a high-quality forensic evaluation in the field of child welfare.

The clinical process involved in conducting a formal forensic psychological evaluations for the child welfare system is often multi-faceted and varies among practitioners. Because there are many types of practices and settings in which forensic evaluations are conducted, quality often varies. Despite the acknowledged importance of forensic evaluations, relatively little empirical research on the quality of reports has been conducted. Little, if any, attention has been given to the manner in which salient information is integrated into reports (Duits, van der Hoorn, Wiznitzer, Wettstein, & de Beurs, 2012; Goodman-Delahunty & Dhami, 2013). It also has not been examined if the order and format of information impact the fact-finding process. It is a common assumption that forensic evaluators expertly integrate relevant information in order to form a conclusion in a report (Goodman-Delahunty & Dhami, 2013; Young, 2016).

Heilbrun (2001) described in his book, *Principles of Forensic Mental Health Assessment*, various practice principles for forensic mental health assessment in general that may be applied across the range of types of forensic evaluations performed. He proposed 29 “common principles” that encompass some of the broad potential issues that may arise. These principles were classified as “emerging” or “established” and were clustered under four broad areas: preparation, data collection, data interpretation, and communication. Heilbrun (2001) constructed these principles with the goals of generalizing training, facilitating research, and developing relevant policies and better practice within the field. Subsequently, he applied these principles to the forensic assessment of sexual offenders and forensic neuropsychological methods (Heilbrun, 2003; Heilbrun, DeMatteo, & Marczyk, 2004). Heilbrun (2001) states in *Principles of Forensic Mental Health Assessment* that the potential for forensic evaluators to produce high-quality forensic assessment for relevant professionals can be enhanced by literature that provides guidance in relevant areas, which is what this current study intends to do. He states that such research can be enhanced into jurisdiction-specific policy, and allow this policy to be empirically informed.

Training on forensic report writing has historically not been guided by research but by legal, ethical, and procedural compliance. This has led to gaps in the quality of forensic assessments and reports (Duits et al., 2012; Fuger et al., 2014). Grisso (2010) discusses discrete types of faults that peer reviewers found in general forensic reports when he analyzed a sample of 62 evaluations that included both criminal and civil forensic issues in the United States. He did not employ a specific criterion or scoring mechanism, but instead asked for “common problems” from the reviewers using a template that offered minimal structure. The problems he identified included opinions being offered without sufficient explanation, forensic purposes

being unclear, organization problems, irrelevant data and opinions, failure to consider alternative hypotheses, inadequate data, data and interpretation mixed together inappropriately, over reliance on a single source of data, language problems, and improper test usage. Forensic psychologists also struggle within their reports to establish connections between clinical data presented and stated conclusions (Wettstein, 2005). Previous research has discovered that another large problem is that forensic practitioners are likely to be satisfied with the quality of their reports and not engage in self-assessment, due to the belief that they do not need to make improvements (Wettstein, 2005, 2010).

Quality Assurance and Improvement in Forensic Practice

Quality improvement and assurance in the field of psychology is important for successful client outcomes (Bonin, 2018). As the current study includes data from a quality assurance study, it is imperative to discuss the literature regarding quality assurance in forensic evaluation. In 2001, Heilbrun wrote that one of the most important needs for research involves “measuring the quality” of forensic evaluations. He purported that quality is not only important for training, practice, and policy, but also to research examining service quality provided to courts and attorneys. Both the quality of the product of a forensic evaluation process and perceived quality by forensic evaluators is important in regard to quality assurance and improvement.

Heilbrun (2001) proposed a model for the practice of forensic mental health assessment based upon ethical, legal, and practice areas, which was then developed into a quality assurance instrument examining assessments of juvenile offenders for a thesis completed by Sanschagrin (2006). Results of her thesis showed that there is a continuing disparity between the theory and practice of forensic mental health assessment. For example, she found that a lot of evaluators did not use psychological testing and opinions relied on interviews and collateral records, even

though that is known to not be best practice. Nguyen et al. (2011) and Fuger et al. (2014), members of the same research team, decided to continue this research in Hawaii examining quality of conditional release reports and quality of criminal responsibility reports using the same instrument. Both studies found that report quality was less than satisfactory, regardless of evaluator professional education or employment status.

Duits et al. (2012) conducted a study of youth forensic mental health evaluations and reports in the Netherlands. This study introduced a quality evaluation instrument called the STAR, which stands for the Standardized Assessment Instrument of Reports. It was developed on the basis of concept mapping to distinguish the different perspectives of usability of the forensic reports. The STAR was created within the Dutch forensic context based on cases in which youths had committed a crime and required a mental health evaluation. For the evaluation, a psychologist was assigned to assess before the trial if the youth had a mental disorder and the impact of the disorder on the crime, if there was a risk for re-offending, and if they had any recommendations for what was best for the youth. The STAR consists of seven clusters or domains created from the following concept mapping: 1) expertise of the evaluator, 2) form of report, 3) history and environment, 4) functioning and development, 5) functioning concerning indictment, 6) advice, and 7) content realization. The STAR enables an objective rating of quality or usability of the mental health report, as well as a score for each domain (Duits et al., 2012). The development of the STAR is one of the only known published instruments used for quality assurance in forensic evaluations. It is similar on that level to the tool used in the current study in that it is an instrument that can support forensic training, peer review, and supervision.

da Silva Guerreiro, Casoni, and Santos (2014) completed a study with the goal of providing a general portrait of forensic psychological reports under the Portuguese justice system. Their study discusses how typically, forensic standards of quality (Melton et al., 2007; Wettstein, 2005) are determined by formal characteristics of these reports, such as organization of content, quality of writing and language used, and the depth of the information conveyed. This allows for critique of lack of depth of writing, poor writing skills, superficiality of language, and lack of organization. However, da Silva Guerreiro et al. (2014) felt as if this was not enough and that examining forensic reports in this manner does not assess the reports' role as informational support for legal decision-making. Therefore, their study examined not only forensic reports in general, but also their "relevance" and "coherence." The study took an example of 106 reports randomly assigned to three data coders. These reports related to two Portuguese laws, Portuguese Criminal Procedural Code's article 159 and article 160. Article 159 referred to forensic psychiatric assessments related to criminal responsibility completed by both a psychiatrist and a psychologist, and article 160 referred to an evaluation that examines "the non-pathological psychological features [and] degree of socialization" of alleged offenders by describing problems of "personality and dangerousness."

A coding "grid" was constructed by da Silva Guerreiro et al. (2014) for both relevance and coherence. Three criteria were used to examine relevance, which was then expanded to further detail with seven elements under these criteria. The study found that of the 106 reports, a substantial number were lacking certain degrees of relevance. For example, some did not contain an assessment of "dangerousness" and some did not contain an assessment of "degrees of socialization." This is important because these topics were directly related to the terminology of the relevant Portuguese law. Similar results were found in both Grisso's (2010) and Lander and

Heilbrun's (2009) studies. Grisso found that the expert evaluator's opinion about the key psycho-legal issue was missing in 56% of the 62 evaluations analyzed. Lander and Heilbrun (2009) examined evaluations completed by primarily psychiatrists in cases of adult criminal misconduct and used a "blue ribbon panel" consisting of a judge, a law professor, an attorney, a psychiatrist, and a psychologist to examine report quality. They found that in about 60% of the evaluations they reviewed ($n=125$), the reasoning explaining the connection between the psycho-legal issues and data collected were not addressed.

da Silva Guerreiro et al. (2014) used two broad criteria to determine report coherence, which were then specified into five elements. Coherence was explored by examining whether the information presented in different sections of the report fit together logically. Results implied that near half of the reports ($n=55$) contained incongruent information, and more importantly, this incongruence was not discussed in the report most of the time. Grisso (2010) also found that when there were alternative explanations possible according to data in a forensic report, 30% of the time they were not discussed. As such, the conclusion was arrived at that forensic reports must clearly discuss sources of data and interpret it appropriately in order to improve report quality and decrease the chance of potential misinterpretation of information (da Silva Guerreiro et al., 2014).

According to Wettstein (2005), in 1996 Dietz was the first to explicitly encourage excellence in result reporting in forensic practice. Dietz stated that forensic practitioners must present findings and opinions with "scrupulous fairness." However, this idea was preliminary to the goal that is more relevant in forensic practice today, which is to address in specific the quality of forensic evaluations. Previous research has been conducted on topics such as the treatment of psychiatric disorders, the use of outcome measures, and consumer ratings of

behavior health services in managed health plans. Yet, research is scarce on the quality of forensic evaluations, and there is little empirical data on quality assurance. Additionally, some opine “quality assurance” has become an umbrella term, and as a result may encompass too many different things and have multiple meanings (Williams, 2016).

Quality assurance is defined by the *Analytic Quality Glossary* (Harvey, 2007) as “the collections of policies, procedures, systems, and practices internal or external to the organization designed to achieve, maintain, and enhance quality.” Williams (2016) asserts that quality assurance can be both an internal and external process and currently the term is used as a type of umbrella term that can have a range of definitions. There has not been a set definition for “quality” in regard to forensic evaluations, even though there have been sets of principles and guidelines put forth by different agencies and authors. Therefore, for the purpose of this study, quality assurance will be defined as stated above.

The Current Study

In summary, despite the growing knowledge base about best practices in forensic evaluations in general, there is a dearth of knowledge about high-quality evaluations in the field of child welfare. Thus, the purpose of the present study is to contribute to the developing knowledge base by describing factors that lead to high-quality forensic psychological evaluations and high-quality recommendations in cases of reported child abuse and neglect. The definition of quality in this study will be defined as an evaluation that adheres to the objectives of New Jersey’s Department of Children and Families’ (DCF) Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (Mental Health) (2012). These particular guidelines and how they relate to each selected predictor will be described within the method section of the current study. The overall goal of this study is to improve future evaluations for the sake of the families

and agencies involved as well as improve the utility of psychological evaluations in child welfare for judicial decision making while simultaneously adding to the literature involving quality improvement in forensic psychology.

Research Questions

With the aforementioned issues in mind, the current study will examine the following research questions investigating predictors of quality in forensic evaluation in child welfare:

1. What are the strongest predictors of overall quality in forensic psychological evaluations in cases involving child abuse and neglect? Predictors will include the presence of a diagnostic assessment in the evaluation, the quality of the evaluator's integration of culture into the evaluation, and clarity of the referral question. This question will be addressed using multinomial logistic regression.
2. Does subject gender or age interact with any of the predictors to affect the overall quality of the forensic evaluation? This will be assessed by testing for interaction effects within the multinomial logistic regression.
3. What are the strongest predictors of the quality of the recommendations in forensic psychological evaluation in cases of abuse and neglect? Predictors will again include the presence of a diagnostic assessment in the evaluation, the quality of the evaluator's integration of culture into the evaluation, and clarity of the referral question. This question will be addressed using binomial logistic regression.
4. Does subject gender or age interact with any of the predictors to affect the quality of the recommendations within the forensic evaluation? This will be assessed by testing for interaction effects within the binomial logistic regression.

Method

Original data utilized in this study were collected by the New Jersey Coordination Center for Child Abuse and Neglect (NJCC). The NJCC, which is a component of the Institute for Families at Rutgers University, was formed in July of 2015 through a partnership with New Jersey's DCF's Division of Child Protection and Permanency (DCPP). The NJCC conducts a study of forensic evaluations conducted in the field of child welfare that are completed by both private practitioners and practitioners who are employed at government sponsored child protection agencies. Its function is to encourage evidence-based practice in child welfare matters throughout the state, thereby improving service delivery and outcomes for children (Rutgers School of Social Work Institute for Families, n.d.). The goal of the NJCC's study is to improve the quality of forensic psychological evaluations in the field of child welfare statewide. The NJCC aims to develop and implement a quality assurance and peer review process within the context of DCF's existing guidelines and best practice. The purpose of the present study, which is conducting a secondary analysis of the NJCC's data, is to explore in depth the factors of diagnoses, cultural competence, referral questions and how they relate to increased quality of forensic psychological evaluations and recommendations in child welfare evaluations. The original study that forms the basis of this research was reviewed and approved by the Institutional Review Board (IRB) of Rutgers University in 2018, 2019, and 2020.

Data collection. The state government divided the state of New Jersey into six catchment areas, and each contain one Regional Diagnostic and Treatment Center (RDTC), which can also be referred to as a Child Protection Center (CPC). These RDTCs were legislatively created to provide expert medical and mental health evaluations and treatment for children who may have been abused or neglected. The centers are also expected to conduct

research into best practices for the forensic evaluation and treatment of child abuse and neglect. The six RDTCs of NJ are the CARES Institute Rowan University School of Osteopathic Medicine (CARES), Dorothy B. Hersch Child Protection Center at The Children's Hospital at St. Peter's University Hospital (St. Peter's), Jersey Shore University Medical Center CPC (Jersey Shore), Metropolitan RDTC Children's Hospital of NJ at Newark Beth Israel Medical Center (Metro), Audrey Hepburn Children's House North Regional Diagnostic Center for Child Abuse and Neglect at Hackensack Meridian Health (Audrey Hepburn), and St. Joseph's Children's Hospital Diagnostic Center for Child Protection (St. Joseph's). Each RDTC encompasses several counties of NJ. Evaluations for the current study were collected from the CARES, St. Peter's, and Metro catchment areas. The CARES catchment area encompasses seven NJ counties, St. Peter's encompasses eight, and Metro encompasses one. Evaluations included in this study were completed by either the RDTC or by private providers contracted with the state.

The current study has sampled 895 ($n=895$) forensic psychological evaluations from the CARES, St. Peter's, and Metro catchment areas. Data were collected from evaluations conducted between 2014 and 2018. Overall, the evaluations sampled by the NJCC account for approximately 15-24% of the total amount of child welfare evaluations completed in NJ in the past four years (Forsythe, 2018). Data from the CARES, St. Peter's, and Metro catchment areas show that 90% of evaluators were psychologists, 3% were licensed clinical social workers, 3% were interns working under the supervision of a psychologist, and 4% were "other" which included masters level clinicians and licensed practicing counselors working under the supervision of a psychologist. There are also cases where an evaluation has been conducted independently by a Licensed Clinical Social Worker or a Licensed Professional Counselor. To

meet inclusion criteria for this study, a provider must have completed at least five evaluations for DCF within a one-year period.

Procedures

The current study used a theory generating approach since there is a lack of literature on the specific factors that impact the quality of forensic evaluations in child welfare. An exploratory approach allows the study to examine all of the relationships in the model despite the lack of literature, in order to provide clear direction of the correlations of the impact on evaluation and recommendation quality. The current study was submitted to the IRB at Rutgers University in 2019 and was approved. The IRB determined that the study did not meet the regulatory determinations of human subject research as it is a secondary analysis of already collected data. It is an exploratory study designed to collect data on factors related to high-quality forensic psychological evaluations and recommendations in the field of child welfare. This study uses a cross-sectional research design and is non-experimental as there is no manipulated independent variable.

Sample

Data were used from 895 forensic psychological evaluations from child maltreatment cases. All evaluations collected from the NJCC that met appropriate criteria for the overall quality improvement study were included in this current study. Evaluations that were considered appropriate for this evaluation included psychological, psychosocial, and parenting evaluations. Evaluations that were not considered appropriate for this study were bonding assessments, fire-setting assessments, and psychiatric assessments.

The subjects of assessment in the evaluations were 22.6% children ($n=202$) and 77.4% adults ($n=693$). The gender of the subjects of assessment were 30.9% male ($n=277$) and 60.6%

female ($n=542$). In 8.2% of cases ($n=73$), the reviewer of the evaluation indicated the evaluator did not specify the subject's gender. In .3% of cases ($n=3$), the evaluator indicated that the subject either identified as transgender or "other." In cases where gender was not specified, each evaluation was examined to infer gender, looking for pronouns and prefixes on the subject of the evaluation's name. Consequently, the gender of the subjects of assessment were 33.4% male ($n=298$) and 66.6% female ($n=593$). Males were coded as "0" and females were coded as "1." Data on ethnicity and language spoken by the subjects were not available for a majority of the evaluations examined and thus could not be evaluated as a factor. Frequency distributions of all study variables as well as crosstabulations are presented in the results section.

Measures

The research questions of this study are based on the Quality Improvement Tool (QI Tool), which is a survey measure that was designed by the NJCC to facilitate a peer review process of forensic psychological evaluations across the state of NJ. The tool is based on DCF's Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (2012). The QI Tool is housed by Qualtrics and has been electronically distributed to peer reviewers who have been contracted by the NJCC as part of the general quality improvement study. The purpose of the development of this standardized instrument was to illustrate a variety of quality improvement possibilities for forensic psychological evaluations. It is designed to gather data in three main areas: case context, key elements of the guidelines, and exploration of practices that may indicate emerging trends in practice. There are thirteen different sections of the QI Tool, with the following titles: Evaluator/Evaluation Details, Referral Information, Case Detail, Background Information, Interview with a Child, Interview with an Adult, Psychological Inventories and Interpretation, Evaluation Questions, Collateral Information, Diagnosis, Summary,

Recommendations, and Evaluation Rubric. The QI Tool is attached in the appendix of this paper.

QI Tool: Code Book. The QI Tool (NJCC, 2018) is based on DCF's Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (2012), but the peer reviewers using the QI Tool are also provided with a Code Book designed by the NJCC. The purpose of the Code Book is to further explain how to answer the questions on the QI Tool. Throughout the Code Book, the DCF Guidelines (2012) are cited within questions so that the user of the book can distinguish which questions map onto specific guidelines. The Code Book lists each question on the QI Tool and the different answer choices. The Code Book also provides definitions for terms commonly used in child welfare within the context of each question, and additionally in a section labeled "glossary of terms." This glossary lists other relevant information as well, such as state laws related to child abuse and terms that may be esoteric to people who work for DCF, so therefore they may be unknown to forensic evaluators. Finally, the Code Book provides references that were used in developing the QI Tool (Darling & Steinberg, 1993; New Jersey Department of Children and Families, 2012; Sullivan, 2011).

Peer-review process. The peer review process used in the NJCC's quality improvement study aims to enhance best practice in forensic psychological evaluations for child welfare by continuously providing feedback on these evaluations. This process is influenced by theories guiding participatory action research (PAR), which is a research methodology based on understanding and then improving something (Baum, MacDougall, & Smith, 2006). PAR differs from most other research approaches in that it is based on reflection, data collection, and action.

The peer reviewers in this study are licensed psychologists with experience conducting forensic psychological evaluations for the state of NJ. They review the work of forensic

evaluators, who are their peers because they complete the same type of work for the state. The NJCC recruited psychologists to serve as peer reviewers and subsequently conducted three peer reviewer training workshops to train the reviewers in using the QI Tool. The workshops included an introduction to the QI Tool and a review of a sample evaluation that was provided to each participant before attending the training. Additionally, peer reviewers have monthly phone conferences with the NJCC staff in order to collaborate with staff, ask questions, and give feedback about the project.

Reliability. There were a number of steps taken by the NJCC to ensure reliability in the use of the QI Tool by the peer reviewers. Once a psychologist completes the training workshop, they review another case to establish reliability with other peer reviewers and accuracy in their use of the tool. In addition, the first ten evaluations they peer review for the project are subjected to automatic secondary review. Thereafter, ten percent of the overall evaluations that are peer reviewed are randomly selected for secondary review by the NJCC staff. Peer reviewers may always request a secondary review if they are not fully confident in their scoring. It is also ensured that the same peer reviewer does not review the same evaluator's work repetitively. All forensic evaluators whose work is reviewed by the peer reviewers using the QI Tool have at least two reviewers scoring their work at some point in time.

The intraclass correlation coefficient (ICC) was used as a measure of reliability. This is a procedure based upon the analysis of variance and the estimation of variable components (Bartko, 1966). As the ICC ranges from 0 to 1, a high ICC that is close to 1 indicates high similarity between values from the same group (McGraw & Wong, 1996). The QI Tool has relatively high reliability as measured by the secondary reviews of evaluations that were

conducted for reliability purposes ($ICC=0.824$). Secondary reliability reviews were conducted by the NJCC for 10.89% ($n=179$) of total evaluations reviewed by the project ($n=1643$).

Outcome Variables

There can be several domains or dimensions of forensic evaluation quality. Relatively little conceptual analysis has been conducted in regard to quality of forensic evaluations (Wettstein, 2005, 2010). However, quality monitoring and improvement of these evaluations have proven to be important issues (Duits et al., 2012). Specific quality indicators for forensic evaluations remain to be developed (da Silva Guerreiro et al., 2014; Grisso, 2010; Wettstein, 2005). Therefore, for this study, two outcome measures of quality were included as separate analyses, overall quality and quality of recommendations. Quality improvement is made possible by measuring defined indicators for performance within evaluations (Duits et al., 2012). Accordingly, despite the lack of information available on quality of forensic evaluations in child welfare, the QI Tool used in this study aims to quantify aspects of quality in order to improve these evaluations.

Overall quality. There is a need to measure quality of forensic evaluations in child welfare, as deficiencies in forensic mental health evaluations currently exist (Combalbert et al., 2014; Grisso, 2010; Wettstein, 2005). Heilbrun (2001) stated that research on quality of forensic evaluations can function as a “crucial” outcome variable for a variety of other factors related to this outcome. Currently, there is a lack of information about quality and how it should be measured in forensic evaluations in child welfare. Defining specific indicators of performance within forensic evaluations is the only way to make quality monitoring plausible (Duits et al., 2012). Therefore, the first outcome measure used in the multinomial logistic regression model in the current study was the overall quality of the forensic evaluation. This number was determined

by the peer reviews conducted using the QI Tool. Quality in the current study can be defined as the standard of an evaluation measured against other evaluations of a similar kind. Peer reviewers were instructed on item 13.22 of the QI Tool to “Please indicate the overall quality of the forensic evaluation written by the evaluator.”

Overall quality of the evaluations was rated by the peer reviewers on a Likert scale ranging from 1-6 points, with “1” being the lowest score and “6” being the highest. The scale is labeled on the QI Tool as “Strongly Disagree” as “1” and “Strongly Agree” as “6”, with no labels for numbers 2-5. Peer reviewers were instructed to give a measure of overall quality based on the evaluation’s adherence to DCF’s Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings. For the evaluations in this study, the mean score on overall quality was 3.71 ($SD=1.26$). The scores given on this item were normally distributed. The variable referring to an evaluation’s overall quality will be labeled “Overall Quality” throughout the results section tables of this study.

Quality of recommendations. Few studies have examined recommendations in psychological evaluations. The literature on recommendations in forensic evaluations typically includes studies that are retrospective court case reviews in custody proceedings examining what variables may have led to specific recommendations (Zumbach, Wetzels, & Koglin, 2018). The quality of recommendations in forensic evaluations is a concept that has not been quantified in a particular way in the current literature. However, there are some studies that have examined the frequency with which psychiatric recommendations are followed by courts, but it is unclear how this relates to forensic psychological evaluation recommendations (Buchanan & Norko, 2013). Research indicates that the comprehensiveness of a recommendation may be related to the quality of the assessment that generated it, but more empirical advances are needed to support

this (Buchanan & Norko, 2013; Zumbach, Wetzels, & Koglin, 2018). For this reason, quality of recommendations was investigated as a separate outcome variable than overall quality of the evaluation.

There are many people involved in forensic cases that stand to benefit from an evaluation's recommendations, including parents, children, service providers, social workers, and legal professionals (Berliner et al., 2015; Pickar & Kaufman, 2013). DCF's Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (2012) stipulate that once the evaluation has been completed by the provider, the summary and report should include relevant recommendations. These recommendations may impact a family greatly, especially if an evaluation was court-ordered, which exemplifies the need for recommendations to be high-quality. Therefore, an evaluator's understanding of what types of recommendations are high-quality may be important in regard to improving outcomes for children and families who are involved in the child welfare system. By using quality of recommendations as an outcome variable, the current study will attempt to measure what factors may be related to high-quality recommendations.

Two items on the QI Tool were averaged to form the outcome variable named "quality of recommendations", as both items measure similar elements as determined by a strong positive correlation ($.595, p < .001$). These two questions state "13.20 To what degree do the recommendations promote the psychological and physical well-being of the child, and when appropriate, facilitate the safe reunification of the child with the parent" and "13.21 To what degree are the recommendations made by the evaluator tied to observable outcomes?" Both of these items were rated by the peer reviewers on a Likert scale ranging from 1-6 points, with "1" being the lowest score and "6" being the highest. The scale is labeled on the QI Tool as

“Strongly Disagree” as “1” and “Strongly Agree” as “6”, with no labels for numbers 2-5.

Recommendations were not provided in 3.4% ($n=30$) of the evaluations. Among the evaluations that included recommendations ($n=865$) the mean combined score of the two questions was 3.95 ($SD=1.12$).

The data for the quality of the recommendations in this study display “sparse data bias” (Greenland, Mansournia, & Altman, 2016). This occurs when it is the situation that if the variable was treated as continuous, the number of events observed would be assumed to be sufficient at all treatment levels to result in appropriate adjusted estimates. But because there is a lack of adequate case numbers for some combinations of predictors and outcome levels, the resulting estimates of the regression coefficients may have bias away from the null. Even though the data include a large sample size ($n=865$), there were very few evaluations that fell into the extreme categories of “1” and “6”. Therefore, a regression analysis could not support inclusion of this variable as continuous due to a large amount of missing data. Consequently, this variable was then dichotomized using a mean split, with evaluations that scored 1-3.95 coded as “0” ($n=333$; 38.5%) and evaluations that scored 3.96-6 coded as “1” ($n=532$; 61.5%). The variable describing quality of recommendations will be labeled “Quality of Recommendations” throughout the results section tables of this study.

There was one other relevant recommendation question that the QI Tool asked the peer reviewers to answer. The question stated, “Does the evaluator include a hierarchy timeline for service-based recommendations to be completed?” Peer reviewers were asked to answer “yes” or “no” to this question. It was found that most evaluations did not include this type of timeline within the recommendations (86.7%, $n=776$). Implications of this will be explored in the discussion section of this study.

Predictors

The selection of predictors for this study was based on forensic literature and followed the theoretical criteria outlined in DCF's Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (2012). The criteria outlined in DCF's guidelines are generally in accordance with the criteria outlined in APA's guidelines for child protection evaluations and the APA's specialty guidelines for forensic psychology as well (2013a, 2013b). All of these sets of guidelines and their relationships to the selected predictors are described below, within the description of each individual predictor and the rationale for its' inclusion in the study. Using chi square analysis, bivariate correlation, multinomial logistic regression analysis, and binomial logistic regression analysis, the following predictors were analyzed in this study: the presence of a diagnostic assessment, the integration of culture into the report, and the clarity of the referral question. Tables of the frequency and descriptive data of the predictors can be found in the results section.

Presence of a diagnostic assessment. Similar to other subspecialties of clinical psychology, forensic psychologists require strong general skills in diagnosis and using formal assessment measures (Mulay et al., 2018). Diagnoses rendered in forensic evaluations have substantial effects on the decisions made by the courts (Gowensmith et al., 2017). Diagnoses provided by evaluators in forensic assessments must be grounded in clinical methods because an invalid diagnosis or formulation may jeopardize the overall validity of the forensic conclusion (Kavanaugh et al., 2006). DCF's Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (2012) state that once a forensic evaluation is completed, the summary and report should include any formal clinical diagnosis arrived at by the evaluator. Budd (2005) stipulates that if diagnostic terms are used within the evaluation, the evaluator should explain what they

mean in lay terms, as well as why they arrived at that diagnosis and how the condition may affect parenting if applicable to the situation.

Presently, there is contention within the field of forensic psychology and psychiatry as to whether or not providing a clinical diagnosis is necessary within a forensic report. The American Association of Psychiatric Law (AAPL) guidelines (Glancy et al., 2015) recommend that when possible, the forensic evaluator should give a formal diagnosis, as it provides a valuable purpose in understanding the evidence or fact at issue. However, forensic practitioners also acknowledge that this use of diagnostic labels may avert attention away from the purpose and function of these evaluations, which typically focuses on individual functioning or parenting abilities of the subjects. Accordingly, the APA Specialty Guidelines for Forensic Psychology (2013b) state that practitioners must consider the problems that may arise when using a clinical diagnosis in some forensic contexts, and qualify their opinions appropriately.

The American Psychiatric Association (2013a) included a special section in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-V) describing use of the manual in forensic contexts. It is a cautionary statement that warns that it has become apparent that even minor changes in the wording in the DSM-V can have unanticipated legal and administrative consequences. For example, there have been updates to diagnoses involving trauma, such as Post Traumatic Stress Disorder (PTSD), that are particularly relevant in the forensic field (Wills & Gold, 2014). Consequently, the DSM-V is described as “not a resource designed to meet ‘all of the technical’ needs” of the legal system (American Psychiatric Association, 2013a). This statement is a generic warning, and forensic evaluators must consider what benefits and risks arise from formulating diagnoses, while also balancing the strengths and

limitations of the DSM-V in the fields of forensic psychology and psychiatry (Wills & Gold, 2014).

Due to the debate in the field as to whether or not conducting a diagnostic assessment as part of a forensic evaluation is beneficial to the overall quality of the report, this was chosen to be a predictor. A diagnostic assessment is defined as a process in which a diagnosis is either given or not given and the diagnostic process is denoted or discussed in some way. In the data included in the current study, subjects sometimes received multiple diagnoses within their evaluation. The average number of diagnoses given was 2.56 ($SD=1.7$), with a range of zero to ten diagnoses given per evaluation. The evaluations were coded into a binary variable using “0” or “1” in relation to whether or not there was a presence of a diagnostic assessment within the evaluation. Evaluations were coded “0” if no mention of an assessment of the subject regarding psychological diagnoses was discussed. Evaluations were coded “1” if the evaluator gave the subject one or more diagnoses, either by history or by the evaluator, or if the evaluator wrote that the subject did not meet criteria for any DSM-V diagnoses. A diagnostic assessment was conducted in 77.8% of cases ($n=696$). Additionally, in 65.3% of cases ($n=584$) a diagnosis was given to the subject of the evaluation, while 34.7% ($n=311$) of subjects did not receive a diagnosis. Peer reviewers were also asked to indicate if there was enough evidence in the report to arrive at the diagnosis given. In the 584 evaluations that contained a diagnosis, peer reviewers rated that 25.8% ($n=151$) of them did not contain enough evidence presented to arrive at the diagnostic conclusions the evaluator described.

In the 696 cases reviewed where a diagnostic assessment was conducted, a diagnosis was provided 84% of the time, with an average of 2.56 diagnoses reported per subject. There were specific diagnostic themes that emerged from the 584 evaluations where a diagnosis was

provided. The most frequently identified diagnoses fell into the following categories of the DSM-V (American Psychiatric Association, 2013b): Trauma and Stressor Related Disorders (41%), Substance-Related and Addictive Disorders (39%), Personality Disorders (30%), Depressive Disorders (28%), Anxiety Disorders (22%), Neurodevelopmental Disorders (22%), Bipolar and Related Disorders (14%), Diagnosis of Child Physical/Sexual Abuse or Child Neglect (10%), Disruptive, Impulse-Control, and Conduct Disorders (8%), Schizophrenic Spectrum and Other Psychotic Disorders (3%), and Other (3%). The Other category consisted of low frequency diagnoses in the following categories: Obsessive Compulsive and Related Disorders, Somatic Symptom and Related Disorders, Feeding and Eating Disorders, Elimination Disorders, Sleep-Wake Disorders, Gender Dysphoria, and Neurocognitive Disorders. Additionally, the peer reviewers were asked to respond “yes” or “no” to the question “Was there reasonably sufficient evidence or basis for the diagnosis made?” About three quarters of the evaluations (74.5%) were rated as having sufficient evidence to support the diagnoses that were given by the evaluators. The variable label used for the presence of a diagnostic assessment throughout the results section tables is “Dx Assessment.”

Integration of culture. A key aspect of any psychological evaluation is cultural competence. Forensic evaluators should be experienced in considering general probabilities and alternative hypotheses during evaluations without profiling (Hicks, 2004). APA’s (2013b) guidelines for forensic psychology emphasize that forensic practitioners must recognize how their own culture, attitudes, values, beliefs, opinions, and biases may affect their professional conduct and impartiality. Research trends indicate that psychological assessment subjects in general may receive improper diagnoses and treatment recommendations if evaluators do not pay attention to ethnic background and context (Alegría et al., 2008; Glancy et al., 2015). Evaluators

must take precautions to not assume the relevance of their cultural knowledge with individual subjects, as culture must be understood from the distinctive perspective of each subject (Aggarwal, 2012). There is also an issue of selecting valid testing measures for different cultures. The AAPL guidelines (Glancy et al., 2015) recommend that forensic evaluators consider how culture informs behavioral problems and state that evaluators should consider culture as part of the case formulation process.

Recently, there has been an emerging debate about the place of culture within the legal system. There is controversy as to whether using culture as a defense in criminal cases is beneficial or harmful. For example, proponents of using culture as a defense state that it is intrinsically unfair to judge someone by rules and values of a society that they do not understand (Kirmayer, Rosseau, & Lashley, 2007). Culture can shape a person's identity by influencing their emotional responses, sense of self, and patterns of thinking and problem solving. However, critics of using culture as a defense state that it can undermine the fairness of the legal system by allowing arbitrary and inconsistent decisions to be made due to individual differences. This dispute is especially important in forensic evaluations in the field of child welfare in the United States. A person coming to the United States from another country may not understand the cultural context of child abuse and neglect within each state which can lead to legal difficulties. However, this is why it is recommended that forensic evaluators should describe cultural context, yet try to steer away from using it as a defense (Boehnlein et al., 2005).

Describing the cultural context of behavior from an outside perspective is not an easy task, especially for the objective forensic evaluator. It is quite necessary though, as it can allow the legal systems involved to reconstruct the logic that may have occurred in the defendant's world (Kirmayer et al., 2007). Framing behavior in this way is a delicate process, as it can give

more information about historical and contextual origins, but can also divide and segregate groups of people. A forensic evaluator must be careful to provide information without stereotyping and stigmatizing communities and group members (Hicks, 2004; Kirmayer et al., 2007). As cultural context can add pertinent information to an evaluation, it is hypothesized that the overall quality of the evaluation will be higher the more culture is properly integrated into the evaluation.

The evaluator's integration of culture into the forensic assessment was rated by the peer-reviewers on a Likert scale ranging from 1-6 points, with "1" being the lowest score and "6" being the highest. The scale is labeled on the QI Tool as "Strongly Disagree" as "1" and "Strongly Agree" as "6", with no labels for numbers 2-5. Peer-reviewers were instructed by the QI Tool on question 13.15 to rate "To what degree does the evaluator take into account the cultural norms of the child/adult being evaluated." For the evaluations in the current study, the mean score on integration of culture was 1.47 ($SD=.94$). The results of this item were positively skewed, as 72.8% ($n=652$) of evaluations received a score of "1", indicating a low degree of cultural integration. The data for integration of culture in this study display "sparse data bias" as well, due to the positive skew and scarcity of evaluations rated highly on this scale (Greenland, Mansournia, & Altman, 2016). Therefore, since there was a lack of adequate case numbers for some combinations of this predictor and outcome levels, the variable of culture was dichotomized with evaluations rated 1 coded as "0" ($n=654$; 73.1%) and evaluations rated as 2-6 coded as "1" ($n=229$; 25.6%). The variable label for integration of culture throughout the results section tables is "Integration of Culture."

Clarity of referral question. Grisso (2010) conducted a review of 62 civil and criminal forensic evaluations, in which he had reviewers critique evaluations and then subsequently

converted these critiques into prescriptive statements. In regard to referral questions, the reviewers stated that the specific questions that the evaluator was asked to focus on should be stated, along with the legal standard that defines the forensic purpose of the evaluation. Additionally, a study conducted by da Silva Guerreiro et al. (2014) hypothesizes that coherence and relevance may be indicators of quality in forensic psychological reports. If reports are not clear about what they are assessing, this can lead to a failure to present information in a logical way. For this reason, the current study examined the clarity of referral questions within forensic evaluations in child welfare.

Common referral questions in child welfare evaluations can address parents' cognitive, emotional, and social functioning; their caregiving skills and deficits; the impact of mental illness and/or substance abuse on parenting ability; risk and protective factors in a family; whether or not child abuse or neglect occurred; characteristics of a parent-child relationship; service recommendations; and service/treatment adherence (APA, 2013a; Budd, 2005). DCF's Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (2012) state that referral questions should determine the specific purpose of the forensic evaluation. One of the identified principles of forensic report writing is that in the first section of a forensic report, the relevant legal questions and forensic issues should be described (Heilbrun et al., 2009; Young, 2016). Conroy (2006) stated that a key factor for good forensic report writing is having an identified forensic reason for referral. Melton et al. (2017) stated that in order to improve forensic reports, evaluators should stay within the scope of the referral question, which implies the referral question should be clearly stated. Grisso (2010) identified frequent report-writing problems in forensic reports and noted that one of the ten most frequent faults in forensic report writing is that the legal question or forensic purpose of the evaluation was not stated, not clear, inaccurate,

or inappropriate. Through synthesis of the current literature and standards that discuss forensic psychological evaluations, it is hypothesized that the clarity of the referral question is integral to the quality of a forensic evaluation in child welfare, therefore this was included as a predictor in this study.

Peer-reviewers were instructed by the QI Tool on question 2.6 to indicate “yes” or “no” to the prompt “Referral questions/statements are clearly stated.” The resulting data from the evaluations were coded into a binary variable using “0” or “1” in relation to whether or not the referral question was clearly stated in the evaluation. Evaluations were coded “0” if the peer reviewer deemed the evaluation did not have a clearly stated referral question. Evaluations were coded “1” if the peer reviewers deemed there was a clearly stated referral question. In 88.4% ($n=791$) of sample evaluations, the peer reviewers deemed the referral questions to be clearly stated, while in 11.6% ($n=104$) cases, the peer reviewers indicated the referral questions were not clearly stated. The variable label for the clarity of the referral question throughout the results section tables is “Referral Clarity.”

Analytic Approach

All analyses were conducted using IBM’s SPSS Version 26. To begin, descriptive statistics and bivariate associations between the criterion variables and predictor variables were examined. Crosstabulations are reported in table format in the results section for model replication purposes. Multicollinearity was assessed by examining the variance inflation factors (VIF) and tolerance measures. These analyses revealed no high levels of multicollinearity in the variables. Next, the data were subject to a series of logistic regression models with the following predictor variables: the presence of a diagnostic assessment, the integration of culture, and the clarity of the referral question. Gender and age category, with adult evaluations coded as “0”

and child evaluations coded as “1”, were inserted into the models to serve as control variables. The first analysis was a multinomial logistic regression with the overall quality of the evaluation as the outcome variable. As the second step, the next analysis added the interaction terms of gender and age category with each predictor variable to the model. The next analysis was a binomial logistic regression with the quality of the recommendations as the outcome variable. As the final step, the next analysis added the interaction terms of gender and age category with each predictor variable to the model. In both logistic regression models, all predictors were added simultaneously, which allowed for examination of the relative importance of each predictor while controlling for the effect of the others. The interaction term analysis added the interaction terms as a second step in both regression models because there was no prior research on these interactions and a theory-generating approach was utilized.

Handling missing values. Missing data analysis described that there were some instances of missing data throughout the data set for certain variables. The outcome variable of quality of recommendations was missing in 30 cases due to these evaluations not including recommendations at all. The integration of culture variable was missing in 12 cases and the diagnostic assessment variable was missing in five cases. Additionally, there were four cases in which gender was not male or female and was instead transgender ($n=1$), other ($n=2$), or not specified ($n=1$). Listwise deletion was used to handle missing data in these cases. Less than five percent of data were missing, which has been previously determined to be an acceptable cutoff to utilize this method of handling missing data (Shaefer, 1999). Additionally, due to the large sample size, there was adequate statistical power to use listwise deletion (Schlomer, Bauman, & Card, 2010).

Results

Descriptive statistics. Descriptive statistics were calculated for each outcome and predictor variable. As seen in Table 1, there was a normal distribution for the outcome variable of quality, with a mean of 3.71, which is why this variable was not collapsed. There were more adult than child evaluations overall. It can be assumed that this is because child welfare agencies are focused on what can be recommended for parents in order to increase child safety. Child evaluations were typically assessments that examined whether or not there was evidence of abuse or neglect. There were also more female than male evaluations. It can be presumed that this may be because there are more female single parents than male single parents, but there is no data from this study to support this assumption at this time.

Table 1

Descriptive Analysis of Frequency Counts and Percentages

		<i>n</i>	%	<i>M</i>	<i>SD</i>
Overall Quality				3.71	1.26
	1	23	2.60%		
	2	162	18.10%		
	3	202	22.60%		
	4	221	24.70%		
	5	240	26.80%		
	6	47	5.30%		
Quality of Recommendations				3.95	1.12
	1-3.95	333	38.50%		
	3.95-6	532	61.50%		
Age Category					
	Adult	693	77.40%		
	Child	202	22.60%		
Gender					
	Male	298	33.40%		
	Female	593	66.60%		
Dx Assessment					
	Yes	696	77.80%		
	No	199	22.20%		
Integration of Culture				1.47	0.94
	1	654	73.10%		
	2-6	229	25.60%		
Referral Clarity					
	Yes	791	88.40%		
	No	104	11.60%		

Correlations. A correlation matrix was computed to ascertain the nature of the association between the non-categorical independent and outcome variables. These correlations were examined before the original variables were recoded, as to measure relationships accurately. Table 2 shows the intercorrelations among variables. Significant correlations between variables were observed but for the most part were within the low to moderate range (0 to .33). The magnitude of these relationships was lower than the commonly accepted multicollinearity threshold of .80 (Munro, 2005). However, the correlation between the two outcome variables, overall quality and quality of recommendations was significant and large.

This positive correlation suggests that as one variable increases, the other increases as well. As both are measures of quality, this makes sense intuitively.

Table 2

Correlations

	1	2	3
1. Overall Quality	-	.738**	.156**
2. Quality of Recommendations		-	-.041
3. Integration of Culture			-

Note: * $p < .05$, ** $p < .001$.

Crosstabulations. Crosstabulations were computed to provide more descriptive information about the predictor variables and their relation to the outcome variables. These calculations took place after the original variables were recoded, as to provide precise information about the variables used in the regression analyses. Tables 3 and 4 show the crosstabulations among variables.

Table 3

Crosstabulations for Overall Quality

		Outcome Variable					
		<u>1</u>		<u>2</u>		<u>3</u>	
		<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Age Category	Adult	14	1.60%	137	15.30%	172	19.20%
	Child	9	1.00%	25	2.80%	30	3.40%
Gender	Male	9	1.00%	51	5.90%	64	7.40%
	Female	14	1.60%	111	12.90%	136	15.80%
Dx Assessment	Yes	16	1.90%	117	13.60%	140	16.30%
	No	7	0.80%	45	5.20%	60	7.00%
Integration of Culture	Lower range	21	2.40%	140	15.90%	135	15.30%
	Higher range	1	0.10%	20	2.30%	65	7.40%
Referral Clarity	Yes	18	2.10%	137	15.80%	182	21.00%
	No	5	0.60%	25	2.90%	20	2.30%

<u>4</u>		<u>5</u>		<u>6</u>	
<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
165	18.40%	175	19.60%	30	3.40%
56	6.30%	65	7.30%	17	1.90%
67	7.80%	87	10.10%	20	2.30%
152	17.70%	153	17.80%	27	3.10%
172	20.00%	210	24.40%	43	5.00%
47	5.50%	29	3.40%	4	0.50%
159	18.00%	169	19.10%	30	3.40%
59	6.70%	67	7.60%	17	1.90%
199	23.00%	209	24.20%	46	5.30%
22	2.50%	31	3.60%	1	0.10%

Table 4

Crosstabulations for Quality of Recommendations Recoded

		<i>Lower quality (1-3.95)</i>		<i>Higher quality (3.96-6)</i>	
		<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>
Age Category	Adult	288	33.3%	378	43.7%
	Child	45	5.2%	154	17.8%
Gender	Male	100	11.6%	188	21.8%
	Female	232	25.9%	341	39.6%
Dx Assessment	Yes	93	10.8%	436	50.7%
	No	94	10.9%	237	27.6%
Integration of Culture	Lower range	240	28.1%	387	45.4%
	Higher range	90	10.6%	136	15.9%
Referral Clarity	Yes	293	33.9%	468	54.1%
	No	40	4.6%	64	7.4%

Multinomial logistic regression analysis. Multinomial logistic regression was used to predict the likelihood of overall evaluation quality based upon the presence of a diagnostic assessment in the evaluation, the quality of the evaluator's integration of culture into the evaluation, and clarity of the referral question. The final model was significantly different ($\chi^2=103.90$, $p < .001$; Nagelkerke $r^2=.117$) from the null model. Table 5 presents the results of the multinomial regression model with the odds ratio (OR) and confidence interval. The OR associated with each predictor and the 95% confidence intervals for each OR represent the effect of an individual predictor (e.g., the degree of cultural integration) on the dependent variable (e.g., the overall quality of the evaluation on the six-point scale). For example, if an OR is larger than 1.00, it depicts the increase of the chance of receiving a specific quality score on the six-point scale. If the OR for a predictor is statistically different from 1.00, it can be observed that the

95% confidence interval does not contain 1.00. The confidence interval can be used to estimate the precision of an OR, with a large confidence indicating a low level of the precision and a small confidence interval indicating a higher level of the precision. However, just because an odds ratio is not statistically significant does not mean that there is no relevant information being provided. The 95% confidence interval does not indicate statistical significance, and the odds ratio can still provide valuable information about an association between the predictor and the outcome even if it is not significant (Szumilas, 2010).

Table 6 depicts each predictor's unique contributions in the regression. The presence of a diagnostic assessment ($\chi^2=30.54, p < .001$), integration of culture ($\chi^2=32.85, p < .001$), clarity of the referral question ($\chi^2=12.80, p < .05$), and age category ($\chi^2=22.81, p < .001$) were significant predictors of overall quality. Gender was a non-significant predictor at every level of the dependent variable. In the second step where the interaction terms were added to the model, it was found that the interaction terms were not significant predictors.

Table 5

Multinomial Logistic Regression Results with Odds Ratios

	Outcome Variable			
	1		2	
	OR	95% CI	OR	95% CI
Dx Assessment	3.83**	[0.94, 15.59]	3.82*	[1.28, 11.41]
Integration of Culture	12.87**	[1.58, 105.12]	3.93**	[1.81, 8.52]
Clarity of Referral	12.81*	[1.36, 120.46]	10.03*	[1.30, 77.42]
Age Category	0.85	[0.29, 2.51]	3.04*	[1.41, 6.55]
Gender	0.92	[0.32, 2.64]	0.71	[.354, 1.42]

Note: OR = Odds Ratio, CI= Confidence Interval; ORs are in reference to evaluations rated “6”, the highest indicator of overall quality of an evaluation and the reference category for the covariates.

Note: Age category reference group is adult (0); Gender reference group is male (0).

Note: * $p < .05$, ** $p < .001$.

3		4		5	
OR	95% CI	OR	95% CI	OR	95% CI
4.4*	[1.50, 12.89]	2.85	[0.97, 8.38]	1.37	[0.46, 4.13]
1.14	[.577, 2.26]	1.48	[0.75, 2.91]	1.43	[0.73, 2.78]
5.67*	[0.73, 44.01]	5.45	[0.71, 41.86]	7.7*	[1.02, 58.28]
3.44**	[1.64, 7.24]	1.58	[0.79, 3.16]	1.66	[0.83, 3.29]
0.73	[0.37, 1.43]	0.64	[0.33, 1.25]	0.86	[0.45, 1.66]

Table 6

Predictors' Unique Contributions in the Multinomial Logistic Regression

Predictor	χ^2	df	p
Dx Assessment	30.54	5	< .001**
Integration of Culture	32.85	5	< .001**
Clarity of Referral	12.80	5	.025*
Age Category	22.81	5	< .001**
Gender	3.26	5	.660

Step 2

Interactions

Age Category x Dx Assessment	3.81	5	.577
Age Category x Integration of Culture	3.76	5	.584
Age Category x Clarity of Referral	4.84	5	.436
Gender x Dx Assessment	5.10	5	.403
Gender x Integration of Culture	4.32	5	.505
Gender x Clarity of Referral	3.41	5	.638

Note: χ^2 = amount by which -2 log likelihood increases when predictor is removed from the full model.

Note: * $p < .05$, ** $p < .001$.

Relationship between the presence of a diagnostic assessment and overall quality.

The results from this study indicated that the presence of a diagnostic assessment within a forensic evaluation in child welfare is a significant predictor of assessment quality. This variable uniquely contributed to the multinomial logistic regression ($\chi^2=30.54$, $p < .001$). For data analysis, evaluations were coded "0" if no mention of an assessment of the subject regarding psychological diagnoses was discussed. Evaluations were coded "1" if the evaluator gave the subject one or more diagnoses, either by history or by the evaluator, or if the evaluator wrote that the subject did not meet criteria for any DSM-5 diagnoses. When examining the odds ratios associated with diagnostic assessment, it can be noted that if the evaluation contained a diagnosis or was coded a "1" it was more likely to obtain a higher quality score. Evaluations without a diagnostic assessment were found to be more likely to obtain a lower quality score.

Relationship between integration of culture and overall quality. Integration of culture ($\chi^2=32.85$, $p < .001$) proved to be a significant predictor of overall quality. The evaluator's integration of culture into the forensic assessment was rated by the peer-reviewers on a Likert scale ranging from 1-6 points, with "1" being the lowest score and "6" being the highest. However, since there was a lack of adequate case numbers for some combinations of this predictor and outcome levels and the data displayed sparse data bias, the variable of culture was dichotomized with evaluations rated 1 coded as "0" ($n=654$; 73.1%) and evaluations coded as 2-6 coded as "1" ($n=229$; 25.6%). By examining the odds ratios associated with cultural integration and overall quality, it can be determined that if the evaluation was coded a "1" it was more likely to obtain a higher quality score. Evaluations that were coded "0" were found to be more likely to obtain a lower quality score.

Relationship between the clarity of the referral question and overall quality. The clarity of the referral question in the evaluations was rated by the peer-reviewers by a response of "yes" or "no" to the prompt "Referral questions/statements are clearly stated." The resulting data from the evaluations were coded into a binary variable using "0" or "1" in relation to whether or not the referral question was clearly stated in the evaluation. Evaluations were coded "0" if the peer reviewer deemed the evaluation did not have a clearly stated referral question. Evaluations were coded "1" if the evaluator deemed there was a clearly stated referral question. In 88.4% ($n=791$) of evaluations, the peer reviewer deemed the referral questions to be clearly stated, while in 11.6% ($n=104$) cases, the peer reviewed indicated the referral questions were not clearly stated. The results of this study revealed that clarity of the referral question ($\chi^2=12.80$, $p < .05$) was a significant predictor of overall quality. Upon examination of the odds ratios associated with the clarity of the referral question and overall quality, it can be determined that if

the evaluation was coded a “1” it was more likely to obtain a higher quality score. Evaluations that were coded “0” were found to be more likely to obtain a lower quality score. This means that if evaluations had clear referral questions, they were more likely to obtain higher overall quality scores.

Relationship between age category and overall quality. The final significant predictor of overall quality in this study was age category ($\chi^2=22.81$, $p < .001$). The two age categories in this study were adult, which was coded “0” and child, which was coded “1.” There were 202 (22.6%) child evaluations and 693 (77.4%) adult evaluations included in the sample. Analysis of the odds ratios appears to suggest that child evaluations were more likely to obtain a higher quality score than adult evaluations. There were no other significant results regarding demographics and the overall quality of the evaluations.

Binomial logistic regression analysis. Binomial logistic regression was used to predict the likelihood of recommendation quality based upon the presence of a diagnostic assessment, the integration of culture into the evaluation, and the clarity of the referral question. The final model was significantly different ($\chi^2=46.60$, $p < .001$; Nagelkerke $r^2=.073$) from the null model. Table 7 presents the results of the binomial regression model with the odds ratio (OR) and confidence interval. The presence of a diagnostic assessment ($\chi^2=14.63$, $p < .001$) and age category ($\chi^2=28.73$, $p < .001$) were significant predictors. Integration of culture, clarity of the referral question, and gender were non-significant predictors at every level of the dependent variable. A second analysis was conducted in which the interaction terms were added to the model in a stepwise binomial logistic regression analysis. It was found that the interaction terms were not significant predictors. Table 8 depicts each predictor’s unique contributions to the regression model.

Table 7

Binomial Logistic Regression Results with Odds Ratios

	Outcome Variable	
	1-3.95	
	OR	95% CI
Dx Assessment	1.94*	[1.38, 2.73]
Integration of Culture	0.90	[0.65, 1.25]
Clarity of Referral	1.06	[0.68, 1.66]
Age Category	2.71**	[1.84, 3.97]
Gender	0.87	[0.64, 1.19]

Note: OR = Odds Ratio, CI= Confidence Interval; ORs are in reference to evaluations rated 3.96-6, the higher indicators of quality of recommendations within an evaluation and the reference category for the covariates.

Note: Age category reference group is adult (0); Gender reference group is male (0)

Note: * $p < .05$, ** $p < .001$.

Table 8

Predictors' Unique Contributions in the Binomial Logistic Regression

Predictor	χ^2	df	p
Dx Assessment	14.63	1	< .001**
Integration of Culture	0.38	1	.537
Clarity of Referral	0.08	1	.783
Age Category	28.73	1	< .001**
Gender	0.74	1	.390

Step 2

Interactions

Age Category x Dx Assessment	0.42	1	.513
Age Category x Integration of Culture	0.12	1	.735
Age Category x Clarity of Referral	0.37	1	.541
Gender x Dx Assessment	3.13	1	.077
Gender x Culture	0.01	1	.975
Gender x Clarity of Referral	0.24	1	.623

Note: χ^2 = amount by which -2 log likelihood increases when predictor is removed from the full model.

Note: * $p < .05$, ** $p < .001$.

Relationship between the presence of a diagnostic assessment and quality of recommendations. The results from this study indicated that the presence of a diagnostic assessment within a forensic evaluation is a significant predictor of the quality of

recommendations of that evaluation. This variable uniquely contributed to the multinomial logistic regression ($\chi^2=14.63$, $p < .001$). When examining the odds ratios associated with diagnostic assessment, it can be noted that if the evaluation contained a diagnosis or was coded a “1” it was more likely to obtain a higher recommendation quality score. Evaluations without a diagnostic assessment were found to be more likely to obtain a lower recommendation quality score.

Relationship between age category and quality of recommendations. The binomial logistic regression analysis revealed that age category ($\chi^2=28.73$, $p < .001$) was a significant predictor of quality of recommendations. Analysis of the odds ratios appears to suggest that child evaluations were more likely to obtain a higher recommendation quality score than adult evaluations. There were no other significant results regarding demographics and the quality of recommendations.

Discussion

The purpose of this exploratory study was to examine possible predictors of quality in forensic psychological evaluations conducted in cases of child abuse and neglect while adding to the quality improvement literature within forensic psychology. Many authors have determined there is a need for quality improvement of forensic assessments (Combalbert et al., 2014; da Silva Guerreiro et al., 2014; Grisso, 2010; Wettstein, 2005). There have been very few studies conducted that use quality improvement tools to measure forensic evaluation quality. One study, conducted in the Netherlands, used a quality improvement tool to systematically examine a sample of forensic evaluations (Duits et al., 2012). This instrument, the STAR, enabled an objective rating of quality or usability of a mental health report, but did not examine which factors led to higher or lower quality reports. A series of studies completed by Nguyen et al.

(2011), Robinson & Acklin (2010), and Fuger et al. (2014) used a quality improvement tool (Sanschagrin, 2006) to examine conditional release reports, competency to stand trial reports, and criminal responsibility reports in Hawaii. Additionally, there has been one other known study that examined the factors of “relevance” and “coherence” and discussed them in relation to forensic report quality in Portugal (da Silva Guerreiro et al., 2014). This research also questioned what valid measures of quality of forensic evaluations should look like. The current study is the first known study examining possible predictors of the overall quality of forensic psychological evaluations and possible predictors of the quality of recommendations in forensic evaluations conducted in cases of child abuse and neglect.

Several research questions were used to guide the purpose of this study. The first research question examined possible predictors of overall quality of a forensic psychological evaluation in cases of child abuse and neglect. The second research question aimed to investigate if the age or the gender of the subject interacted with any of these predictors. The third research question examined another measure of quality, the quality of recommendations, and its possible predictors, in the forensic psychological evaluations. The final question examined whether the age or the gender of the subject interacted with any of the predictors as well.

The variables in this study were measured by the QI Tool, an instrument designed by the NJCC that supports forensic evaluation training, peer review, and supervision. A series of logistic regression models were analyzed with the following predictor variables: the presence of a diagnostic assessment, the integration of culture, and the clarity of the referral question. It was found that the presence of a diagnostic assessment, integration of culture, the clarity of the referral question, and subject age category were significant predictors of overall quality.

Additionally, it was found that the interaction terms were not significant predictors of overall quality. Results also showed that the presence of a diagnostic assessment and age category were significant predictors of quality of recommendations. Again, it was found that the interaction terms were not significant predictors of quality of recommendations.

Relationship between the presence of a diagnostic assessment and overall quality.

The results from this study indicate that the presence of a diagnostic assessment within a forensic evaluation in child welfare is a significant predictor of assessment quality. As stated in the results, this variable uniquely contributed to the multinomial logistic regression. This variable's contribution to the overall regression is interesting because there is disagreement in the field about whether or not conducting a diagnostic assessment as part of a forensic evaluation is beneficial to the overall quality of the report. Some forensic evaluators opine that stigma associated with diagnostic labels may distract from the purpose and outcome of the evaluation (Kavanaugh et al., 2006), which is typically to address an individual's functioning and/or parenting or to state evidence in support of or not in support of whether child abuse/neglect has occurred (Budd, 2005). Stigma in general associated with diagnostic labels has been proven to interfere with clients' willingness to seek care and with adequate provision of care (Craddock & Mynors-Wallis, 2014). However, when a diagnosis is provided within an evaluation it can be expected that there is more information surrounding this diagnosis within the case formulation, as it is well-known within the field that a diagnosis alone is not sufficient (Budd, 2005; Craddock & Mynors-Wallis, 2014). Wortzel (2013) proposes that forensic evaluators may even be able to enhance their credibility and the court's faith in their expert opinion by illustrating how nuances in diagnosis do not change legally defined constructs.

Data analysis revealed that not only was the presence of a diagnostic assessment a significant predictor of quality, but also that in evaluations where there was a diagnostic assessment, the overall quality was more likely to be higher. It is hypothesized that this occurred due to the likelihood of there being more information in the report when a diagnosis was provided. Ethically, forensic evaluators should be using case-specific evidence when assessing and describing a subject's functional and conditional abilities in relation to their symptoms and demonstrated capacities (Young, 2016). It can be inferred that this means that if a diagnosis is provided, the evaluator should be relating this diagnosis to the subject's abilities and capacities, thus leading to more information being provided than there would be in a report in which diagnosis is not included.

Forensic psychological evaluations should distinguish between diagnosis, impairment, and disability (Gold, 2013). In forensic evaluations in child welfare, descriptions of symptoms and diagnoses have potential to bias the readers of the report, which is why some evaluators do not conduct diagnostic assessments (Wills & Gold, 2014). However, this ignores the fact that sometimes diagnostic assessments and symptom descriptions may be helpful in obtaining more services for the family. For example, if an evaluator discovers that a parent who is being investigated by child protective services for abuse or neglect is cognitively impaired, removal of the children does not always have to be the first option. Services such as an in-home therapist and parent coach can be implemented to train the parent in basic parenting skills. This can be recommended by the evaluator if appropriate, and the diagnosis may actually help the parent qualify for more state-sponsored services as well.

It is notable to discuss the range of the number of diagnoses that were provided within the forensic evaluations in the sample of the current study. The average number of diagnoses given

was 2.56 ($SD=1.7$), with a range of zero to ten diagnoses given per evaluation. This means that at least one evaluation contained ten diagnoses of one subject. As researchers have currently called for the need to improve diagnostic accuracy (Alegria, Vallas, & Pumariega, 2010; Liang, Matheson, & Douglas, 2016) evaluators need to ensure they are not misdiagnosing or over diagnosing subjects. This could lead to inappropriate and ineffective care. This is especially pertinent when working with individuals from racial, ethnic, cultural, and linguistic minority backgrounds, as research continues to indicate that this population is receiving poorer quality of care for their mental health difficulties as compared to their non-Hispanic Caucasian counterparts (Liang et al., 2016).

Relationship between integration of culture and overall quality. Across the United States, overrepresentation of people who are linguistically and culturally diverse within the child welfare system is well documented (Svevo-Cianci & Lee, 2010). Understanding where disparities occur within the child welfare pathway is currently a topic that is not well understood (Dettlaff et al., 2011). This may point to bias within the child welfare system, and forensic evaluators must make sure they are addressing this possibility within their evaluations. The level of stigmatization of mental health diagnoses and symptoms may dictate how people of different cultures express themselves (Liang et al., 2016). This may be further altered within the context of a forensic evaluation in child welfare, due to the problems that are the topic of assessment likely existing within the family structure (Dubowitz et al., 2011). Integration of culture was found to be a significant predictor of overall quality. In order to measure the degree of cultural integration in this study, peer reviewers were asked to respond to the following prompt: “To what degree does the evaluator take into account the cultural norms of the child/adult being evaluated.” Nearly three quarters of the evaluations in the current study’s sample were rated the

lowest score on the scale for this prompt. According to the literature, these results are not surprising.

Neal and Brodsky (2016) posit that forensic evaluators are occupationally socialized to view themselves as objective in their practices. However, developing research reveals that this may not be accurate, which presents the need for more studies on mitigating biases (Zapf & Dror, 2017). Reliability improves among evaluators when standardized measures are used, however there are often not enough appropriate measures to use with individuals who are racially, culturally, and linguistically diverse (Gowensmith, Sledd, & Sessarego, 2015). Before conducting psychological assessments, it is recommended that evaluators consult relevant literature to determine whether their test battery is appropriately normed for the individual being evaluated (Chiu, 2014). There is a large literature base that confirms that there is enormous variability across common test instruments such as the MMPI and MCMI regarding cross-cultural and cross-national influences (Archer & Vauter, 2016).

It was found that the evaluations that were rated as integrating culture more were more likely to have higher overall quality scores. It was hypothesized that the overall quality of the evaluation would be higher the more culture is properly integrated into the evaluation, so the results were congruent with the study's expectations. It was predicted that including information about culture in a forensic evaluation in child welfare would increase the amount of information about a subject, thus increasing the report's quality. However, some research suggests that providing too much information about a subject's culture may lead to stereotyping and become detrimental (Kirmayer et al., 2007). Hicks (2004) makes a valid point in stating that cultural background can be very important if it is described within the family context. Variability in the experience of mental illness for different cultures implies that there may be differences in

expression of symptoms of mental illness as well (Liang et al., 2016). For this reason, expert evaluations and opinions in child abuse and termination of parental rights cases must describe cultural issues while simultaneously distinguishing between science, stereotypes, and speculation (Hicks, 2004).

Relationship between the clarity of the referral questions and overall quality. Budd (2005) states that forensic evaluations are not able to answer questions unless they have been articulated properly by the referral source. This highlights the importance of having the reasons for referral communicated clearly to the evaluator prior to the assessment. Grisso (1987) characterized competent forensic psychological assessments as “specific”, meaning that they should efficiently answer the question that was asked by the referral source with necessary and logical clinical and forensic information. For this reason, it intuitively makes sense that referral questions should be clearly stated. The results of this study revealed that clarity of the referral question was a significant predictor of overall quality. Additionally, it was noted that the evaluations that were rated as having clear referral questions were more likely to obtain a higher overall quality score.

Approximately ten percent of the evaluations analyzed in this sample were rated as not having a clear referral question. This percentage seems to be relatively low, and this may be due to the specific type of evaluation of the sample, which was forensic psychological evaluations in child welfare conducted in NJ. The evaluators completing this type of forensic evaluation are trained in the specific Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings put forth by NJ’s DCF (2012). As these guidelines state unambiguously that referral questions should reflect the purpose of the evaluations, it appears to be that most evaluators followed these guidelines and stated why the evaluation was being conducted.

In evaluations completed in the field of child welfare and in forensic evaluations in general, sometimes referral questions can become unclear across involved parties (Budd, 2005). Caseworkers and attorneys who request evaluations may have different reasons for referral and varying expectations about what the report should include. Evaluations can also be a “fishing expedition” with no specific question in mind (Budd, 2005). Different parties interested in the evaluations, such as social workers, child welfare workers, attorneys, law enforcement, and family members may present information related to the referral question that they deem as factual, but the evaluator must consider this information as opinion and then consider the evidence that supports it (Budd, 2005; Zapf & Dror, 2017). In these types of cases, it is also important for the evaluator to remember that their report must be decipherable to the layperson and not solely involved parties who are familiar with the case (Kuehnle et al., 2000). It has been purported that forensic evaluators in general may lose track of the referral reason due to extraneous information provided to them (Zapf & Dror, 2017). Evaluators may receive and review a lot of information about a subject that may not be pertinent to the cases. The most potentially biasing aspect of this information is that it could contain inferences made by others. These inferences may be relevant to the referral question but also may not be relevant to the particular case. Evaluators must remember to be clear with which questions they are addressing in their evaluation and to not be swayed by others’ inferences, as evaluators typically do not know the reasoning process and logic used by interested parties to arrive at their conclusions (Zapf & Dror, 2017).

When Grisso (2010) identified frequent report-writing problems in forensic reports, he noted that one of the most common issues was that the legal question or forensic purpose of the evaluation was not stated, not clear, inaccurate, or inappropriate. Wettstein (2005) also found

that psychologists struggled to establish logical connections between the data in their evaluations and their conclusions. Relating information back to the referral question can help the data remain specific and pertinent to the legal matter at hand. da Silva Guerreiro et al. (2014) suggest that quality of forensic reports can be measured by their relevance, which can relate to both the legal criteria of the evaluations and the degree to which the information included is unique to the individual assessed. It seems that the criteria of “relevance” this article describes would include the factor of whether or not a referral reason is clearly stated within the forensic report. A strong and clear referral question sets the tone for the forensic evaluation, and best practice suggests that the relevant legal questions and forensic issues should be described in all forensic evaluations (Heilbrun et al., 2008; Young, 2016).

Relationship between covariates and overall quality. There were two covariates used in the statistical models for this study, subject age category, which was defined as whether the evaluation subject was an adult or a child, and gender. No other demographic variables were able to be used due to the amount of missing demographic data within the forensic evaluations, such as ethnicity and language spoken. In both logistic regressions conducted in the current study, the subject’s age category was a significant predictor of measures of quality. Furthermore, analysis of the odds ratios revealed that child evaluations were more likely to receive higher scores on the outcome measures of quality.

In NJ, some RDTCs conduct only child evaluations, while others conduct both child and parent evaluations. These centers were created by the state in order to directly provide expert medical and mental health evaluation and treatment for children who may have been abused or neglected. RDTCs also conduct research into best practices for the forensic evaluation and treatment of child abuse and neglect. Therefore, it may be assumed that evaluators conducting

evaluations in this setting may be held to a higher standard, or at least be expected to strictly adhere to DCF's Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (2012). The same cannot be inferred for private contracted providers of forensic evaluations in child welfare. It is hypothesized that since more child evaluations are likely to be conducted by RDTCs, due to some specializing in child evaluations only, that this is the reason that child evaluations were found to generally be of higher quality. Psychologists at RDTCs may be more intensively trained due to being government sponsored agencies that also conduct research. Specific data on which evaluations were conducted by RDTCs versus private contracted providers is not available due to the data used within this study being taken from a de-identified sample. As such, why this occurred within this sample can only be speculated, and should be further explored.

Another possibility as to why child evaluations generally scored higher on both measures of quality within this study is that child evaluations may have contained more information. In clinical psychology, when a child evaluation is conducted it is best practice to interview collateral sources, such as parents, teachers, and any other involved service providers (Mulay et al., 2018). This standard is typically carried over to forensic evaluations, and should be especially practiced when a case is particularly divisive (Herman, 2005; Melton et al., 2017). Children may not always be the most reliable informants about a situation, especially when evaluators need to investigate reported occurrences of abuse or neglect. A child may have been groomed by a perpetrator or not fully educated about abuse and body safety, so forensic evaluators must seek out collateral information in order to write a reputable report (Herman, 2005). DCF's Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (2012) state that evaluators in some circumstances may have to rely on the formal observations

conducted by other neutral and competent professionals. Accordingly, it is hypothesized that in this sample, evaluators conducting child evaluations may have accessed collateral information from other sources, thus making their evaluation more detailed which resulted in higher quality scores. Evaluators conducting evaluations of parents may have had more difficulty contacting collateral sources, or attempted to do so less often.

Relationship between diagnostic assessment and quality of recommendations. DCF's Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (2012) state that evaluations may contain treatment recommendations, but also must primarily function to inform the parties involved and assist the court in rendering decisions in child welfare. The guidelines go on to state that recommendations provided should be intended to reduce further risk to children involved in the case and should focus on rehabilitation for the parent. An additional purpose of the recommendations is described as to recommend interventions that promote the psychological and physical well-being of the child and when appropriate, facilitate the safe reunification of the child with the parent. The binomial logistic regression model conducted with quality of recommendations as the outcome found that the presence of a diagnostic assessment within an evaluation was a significant predictor with unique contributions to the model. Analysis of the statistically significant odds ratio describing the relationship between the presence of a diagnostic assessment and the quality of recommendations revealed that if there was a diagnostic assessment within an evaluation, it was more likely to obtain a higher quality of recommendations score.

Psychological diagnosis that occurs within a clinical psychology evaluation should be part of a formulation that synthesizes subject history, test results, and current functioning. This formulation should typically be followed by recommendations or a management plan for the

subject and other recipients of the evaluation (APA, 2013a; Craddock & Mynors-Wallis, 2014). Forensic evaluations can be structured in the same way when a diagnosis is provided. When a diagnosis is given in a forensic evaluation, evaluators must describe their opinions as to how genuine the subject's symptoms are and determine if the subject's clinical presentation aligns with the legal standard in question (Davis & Lister, 2019). Consequently, it can be hypothesized that when a diagnosis is given, evaluators may provide more information and detail about the clinical formulation they developed, which in turn may lead to more specific, high-quality recommendations.

Relationship between covariates and quality of recommendations. The subject's age category, either child or adult, proved to be a significant predictor of a forensic evaluation's quality of recommendations. As aforementioned, in both logistic regressions conducted in the current study, the subject's age category was a significant predictor of measures of quality. Moreover, analysis of the statistically significant odds ratios from the binomial logistic regression indicated that child evaluations were more likely to receive higher scores on the measure that examined quality of recommendations of an evaluation. It is difficult to estimate exactly why the patterns in this data revealed that child evaluations typically had higher quality of recommendations than adult evaluations. As mentioned previously, child evaluations may have been more detailed, or evaluators may have accessed more collateral material to integrate into the child evaluations. Additionally, it is possible that evaluators who conducted child evaluations may have received intensive training in conducting forensic interviews with children and intensive "on-the-job" supervision and feedback (Herman, 2005).

Specificity of recommendations. It is well known that a recommendation within a psychological evaluation should be specific (APA, 2013a; Berliner et al., 2015; Heilbrun, 2001).

This study examined an aspect of how specific the recommendations in the evaluations were, which was whether or not the recommendations contained a hierarchy timeline for service-based recommendations. Most of the evaluations (86.7%) did not contain a timeline. Forensic psychological evaluators must consider how a family may have difficulty fulfilling many recommendations at once, and that caseworkers and the court may not always have an idea of which services would be appropriate at which time. For example, in a case where reunification is the goal, it may not be appropriate for a victim of child abuse to have visits with a family member before getting individual trauma treatment. Forensic psychologists may be aware of this while other parties may not be. Evaluators must ensure that recommendations are understandable and attainable, especially to individuals that do not have a foundation in psychology (Berliner et al., 2015; Kuehnle, Coulter, & Firestone, 2000). For this reason, it may be important for evaluators to consider the specificity of their recommendations.

The American Professional Society on the Abuse of Children has practice guidelines that speak to evidence-based service planning for child welfare. These guidelines state that service plans should be assessment informed and avoid over-burdening families with compulsory services (Berliner et al., 2015). The emphasis is placed on effectiveness, efficiency, and parsimony, while prioritizing services by sequencing them. For example, if a family is homeless and a parent has substance abuse problems, those needs should most likely be met before requiring parent therapy. Forensic psychological evaluators may benefit from incorporating recommendations into their evaluations by using guidelines set forth in evidence-based service planning for child welfare (Berliner et al., 2015).

Study Impact on Forensic Psychological Evaluations

Forensic evaluations have considerable influence on how information is presented to the court (McCallum & Gowensmith, 2019). Results from this study can add to the literature that exists about forensic psychological evaluations. However, there is not an abundance of literature regarding forensic evaluations in child welfare. Therefore, the results of this exploratory study will be mapped onto studies that examined quality of forensic evaluations in general, with the hope that more studies directly related to child welfare and maltreatment will be completed in the near future.

A comprehensive review of the literature involving forensic evaluations and quality improvement was conducted and yielded three peer-reviewed studies. The studies were conducted in three countries: Portugal, the Netherlands, and the United States. However, the studies conducted in the United States took place in Hawaii, which handles forensic evaluations differently than the rest of the United States. Each country has different laws and jurisdictions, thus rendering procedures and results drastically different. Additionally, the studies described below are not about evaluations conducted in the field of child welfare. Due to the differences between each location and the current study, each study will be described separately in comparison to the current study.

Portuguese forensic psychological reports. da Silva Guerreiro et al. (2014) conducted a study that provided a general portrait of forensic psychological reports completed in Portugal that were conducted with individuals accused of criminal offenses. From conducting a literature review, the authors decided they wanted to measure the concepts of “relevance” and “coherence” within these reports. They hypothesized that these specific characteristics may be related to the overall quality of the forensic reports. A coding “grid” was constructed to measure these two factors within a sample of 106 reports. There were three coders who were asked to reply “yes”

or “no” to the criteria listed on the coding grid. Three criteria were used to examine relevance, which was then expanded to further detail seven elements under these criteria. The three main criteria were, “A clear methodology was employed”, “various sources of information are used and their relative importance is taken into account”, and “assessment goals [associated to articles 159 or 160] are addressed.” Two criteria were used to determine report coherence, which was then specified into five elements. These criteria stated, “Presence of coherent information about the assessee across different sections of the report” and “Explanations for behavior follow from information previously reported.”

In regard to relevance, the study found that of the 106 reports, about half of them used a clear methodology, and about a tenth of them used various sources and discussed their relative importance. Most of the reports addressed the assessment goals related to the Portuguese law in their conclusions, however only 9.4% of evaluations discussed test data and how it related to the individual assessed. It was found that it was rare to see qualitative information about the evaluation’s subject being integrated with quantitative data. In regard to coherence, the study found that roughly a quarter of evaluations presented information coherently across different sections of the report while around 7.5% of evaluations explained behavior from information previously reported in the body of the evaluation (da Silva Guerreiro et al., 2014). It is expected that psychological reports in general communicate what is specific and unique about the individual in not only the Portuguese setting, but around the world (da Silva Guerreiro, 2014; Duits et al., 2012). This may include speaking about demographics and personality in connection to test results. This study found that forensic evaluators failed to do this, and this can be related to the cultural variable in the current study. Evaluators in the current study often did not report demographic information about their subjects, which could have led to a lack in the

“relevance” of the reports. When cultural information was not integrated into the reports, it was found that the report quality was rated as typically lower.

Evaluations conducted in the forensic context should report any inconsistencies noted between their subject’s behavior, test scores, accusations, and personality. It is important that incoherence in information stated is discussed and explanatory hypotheses are noted for a subject’s behavior (Combalbert et al., 2014; Grisso, 2010). In the current study, it was found that if there was a presence of a diagnostic assessment within the evaluation, it was more likely to be rated as having a higher quality and having a higher quality of recommendations. This speaks to the coherence of these evaluations. It can be inferred that these evaluations may have had more information and hypotheses within them due to them providing a diagnosis and likely supporting that diagnosis with information. If information such as this is integrated into evaluations, potential misinterpretations may be prevented, rendering these evaluations increasingly more high-quality and services by psychologists more useful.

da Silva Guerreiro and colleagues (2014) additionally found that most formal elements of the forensic evaluations they examined were present, such as an assessment of personality aspects, dangerousness, and degree of socialization, since these must be stipulated according to Portuguese law. However, the authors make the relevant point that just because formal characteristics are mentioned in a forensic report does not mean that the report is relevant and coherent. This was also found in the current study, as all of the evaluations offered opinions about child welfare cases, yet not all evaluations were rated as having indicators of high-quality. The concepts of “relevance” and “coherence” are not yet valid measures of forensic report quality, but the research completed by da Silva Guerreiro et al. (2014) relates to the current study in many ways.

Forensic mental health evaluations on youth in the Netherlands. Duits et al. (2012) constructed a quality evaluation tool called the STAR to examine forensic mental health reports of youth in the Netherlands. The Dutch forensic context is different than that of the United States and this comparison is interesting to discuss in relation to the current study and for future research possibilities. In their forensic evaluations of youth, Dutch evaluators are required to clarify if and to what extent a mental disorder or deficient mental development has had influence on the offense in question, give an estimate of the risk of re-offending, and give advice about the most favorable development for the subject of the evaluation. However, there are noted organizations responsible for ensuring the quality of these evaluations that are conducted by independent psychologists and psychiatrists. The Dutch Medical Disciplinary Boards stipulate criteria for forensic mental health reports: comprehensiveness, consistency of conclusions, descriptions and justifications of the grounds of conclusions, forensic experts staying within the boundaries of their expertise, and that the method of the diagnostic assessment should lead to answering the questions of the court (Duits et al, 2012). This is different from the United States in that diagnostic assessment is not required in evaluations, and there is debate about whether or not it is appropriate. The current study found that the presence of a diagnostic assessment may be related to higher quality evaluations. The United States requiring diagnostic assessment like the Netherlands may be a worthwhile endeavor, especially since the evaluator can arrive at the conclusion of “no diagnosis” and that still technically is providing a diagnostic assessment.

The Netherlands Institute of Forensic Psychiatry and Psychology (NIFP) is one entity instead of separating psychiatry and psychology, which is not the case in the United States. The NIFP evaluates and improves overall quality of forensic reports by helping to develop referral questions for courts as well as standards and formats for evaluations. It also offers professional

peer review services and organizes forensic training, supervision, and expert meetings (Duits et al., 2012). The current study found that when the referral question within an evaluation was rated as clear, the report often had higher overall quality. This may be related to why the NIFP assists in developing referral questions for the courts.

Although the Netherlands has organizations that are seemingly invested in the quality of forensic reports, there was no standardized instrument to measure evaluation quality. Accordingly, Duits et al. (2012) developed a quality improvement tool, the STAR, based upon “usability” of a report, which they determined to be the most important quality aspect of the report. They used a concept mapping model to develop the STAR, with input from various stakeholder groups. The STAR consists of 46 items within seven domains, 1) expertise of the evaluator, 2) form of the report, 3) understanding and explanation of the subject’s history and environment, 4) understanding and explanation of the subject’s functioning and development, 5) functioning concerning indictment, 6) advice given, and 7) content realization. Each domain is supposed to indicate how usable a report is, and the raters using the STAR rated their answers on a 4-point Likert scale: 1=poorly/no/missing, 2=satisfactory/doubtful/partially, 3=good/yes/present, and 0=not relevant. Subsequently, the researchers transformed the Likert ratings to a 1 to 10 score for the items using a web-based version of the STAR, resulting in a total score for each domain plus a total score for the report (Duits et al., 2012). It was unclear how this was done.

Similar to the current study, the STAR was used to review evaluations conducted previously, except they were evaluations conducted on violent youth offenders instead of evaluations conducted on children and parents involved in the child welfare system. In total, 690 reports were analyzed by the STAR, and the article reports overall domain and overall quality

scores in years 2005, 2006, and 2007. The STAR found that the quality of most of the domains it measured as well as the evaluation overall quality significantly increased from 2005-2007. This was true for each domain except in the area of content realization, which looked at whether reports were consistent, unbiased, and if the evaluator answered the questions asked of them. Additionally, differences between groups were found in regard to different geographical regions of the Netherlands, as well as differences in quality between reports of Dutch and ethnic minority reports. Specifically, reports completed with ethnic minorities as the subjects had significantly lower quality in two domains, the understanding and explanation of the subject's functioning and the subject's development and functioning concerning indictment (Duits et al., 2012). This relates to the findings of the current study in that evaluations were found likely to have lower scores on overall quality when there was a low degree of cultural integration. Overall, the STAR had many differences from the QI Tool, as it was created to for a different country and for a different type of forensic reports. Yet there are many similarities in both the idea of a quality measurement tool and in the results of the analyses yielded from the tools. Both instruments are valuable for research purposes, peer review utilization, and forensic training within the forensic context they were developed.

Quality of forensic reports submitted to Hawaii judiciary. Although Hawaii is a part of the United States, it uses a unique "three-panel" system for assessing trial felony competency to stand trial, criminal responsibility, and post-acquittal conditional release (Acklin, Fuger, & Gowensmith, 2015). This means that three different mental health professionals must independently examine an individual and then relay their opinions to the court in a written report. Typically, these professionals consist of one psychiatrist and two psychologists. Various researchers recently examined the quality and reliability of the mental health reports that were

submitted to the Hawaii judiciary (Acklin et al., 2015; Fuger et al., 2014; Nguyen et al., 2011; Robinson & Acklin, 2010). These studies all modified an objective survey instrument designed for evaluation of juvenile forensic assessments by Sanschagrín (2006). This survey has 44 items of which yield a Quality Coefficient (QC) score. The QC was calculated by dividing the total score of each report by the maximum possible score, which was then converted to a percentage score. Report quality criterion was set at 80% in three different studies examining criminal responsibility, conditional release reports, and competency to stand trial reports, respectively (Fuger et al., 2014; Nguyen et al., 2011, Robinson & Acklin, 2010). All of these studies found that it was rare for an assessment to meet the quality criterion of 80%. They also arrived at the conclusion that there is a less than satisfactory agreement between the panel of evaluators who complete these forensic assessments (Fuger et al., 2014).

Sanschagrín (2006) developed a survey instrument for her doctoral dissertation that included essential and recommended components of forensic mental health assessments broadly based upon Heilbrun's (2001) principles of mental health assessment. On this 44-item instrument, three different areas are examined including identification, legal, and clinical elements of the assessment. Information for each component is coded for its presence or absence and then separately for its quality. Overall quality was represented by the total score, which was coded as 0=absent, 1=present but insufficient, and 2=sufficient for each item. The study examined 180 juvenile mental health assessments completed by 16 different evaluators (Sanschagrín, 2006).

Sanschagrín's (2006) quality measurement tool has definitely contributed much to the field in regard to examination of the inclusion of certain elements in forensic assessments. There are a lot of similarities as well as differences between this measure and the one used in this study.

One of the similarities is that the tool asks, “Did the evaluator identify the legal question being addressed?” This question is very similar to the predictor included in the current study, which asked if the referral question was clearly stated. Of the 180 reports in the Sanschagrín (2006) study, only 6.1% of evaluators were rated as completely identifying the legal question being addressed, with 62% partially addressing the legal question. The current study shows a marked improvement in this, as a majority of the evaluations were rated as clearly stating the referral question.

The measure Sanschagrín (2006) created does not examine the presence of a diagnostic assessment within the assessment as thoroughly as it is discussed in the QI Tool. Diagnosis is mentioned once within the survey tool codebook, and not once within the actual quality measure. This is interesting as this variable was included as a predictor in this study due to the significant impact it can have upon the case, especially in the legal arena (Davis & Lister, 2019; Craddock & Mynors-Wallis, 2014). Furthermore, one of the most intriguing differences between the tool Sanschagrín (2006) developed and the QI Tool of the current study is the mention of culture. The QI Tool asks the peer reviewer to rate how well culture was integrated into the report on a Likert scale, while the Sanschagrín (2006) quality measurement tool does not mention culture at all. Inclusion of cultural aspects into a forensic evaluation, whether psychiatric or psychological, has proven to be critical to gain a thorough understanding of the subject and their behaviors, so this is a surprising domain to exclude from a quality measurement tool (Aggarwal, 2012; Alegria et al., 2008; Allan & Grisso, 2014). This may have occurred due to the impact of culture not being as well known or researched in the forensic context in 2006 as opposed to 2019 (Boehnlein, Schaefer, & Bloom, 2005; Chiu, 2014). Even presently, the current study found a

lot of evaluators did not mention important cultural aspects of the individual being evaluated, so this is still an area of forensic assessment that requires improvement.

Limitations of the Study

The results of the current study have implications for the impact of several predictors on measures of quality in forensic evaluations conducted in child welfare. These results can provide forensic evaluators with important evidence to inform best practices in forensic evaluation. However, the results are entirely based upon one method of data collection, the QI Tool. Thus, the problem of rater bias and mono-method bias are limitations of the study. Additionally, the QI Tool has some weaknesses in its design as well. The Likert scale for the items is labeled “Strongly Disagree” for choice “1” and “Strongly Agree” for choice “6” on the scale. The labels of this scale do not always make sense with the question asked, which was an error in construction of the scale. However, the QI Tool has been found to be reliable across different raters, and is one of the first tools of its kind used among evaluators to aid in peer review of fellow evaluators’ work. It is a promising start to evaluating measures of quality in forensic evaluations even with its noted limitations.

There were not enough evaluations that described the subject’s demographic and cultural data for this information to be included in this study. Evaluators did not mention subject race, ethnicity, or language spoken in a vast majority of the evaluations. Therefore, quality of evaluation could not be compared between these categories, thus implicating a loss of potentially valuable information. It can be argued that an understanding of a subject’s culture can lead to a better understanding of a subject’s origins of behavior and the level of intent and volition a subject had when engaging in a behavior (Kirmayer, Rousseau, & Lashley, 2007). This is

especially relevant in cases of child abuse and neglect because some cultures view physical discipline through a different lens than that of state laws in the United States. Forensic evaluators and other involved parties may take an interest in whether or not they are judging subjects exclusively by the laws, values, and rules of a society of which they are unfamiliar (Lansford et al., 2015). Cultural factors may change the recommendations and diagnoses that a forensic evaluator arrives at during the evaluation process. Evaluators should examine the impact of cultural factors relevant to the conclusions and recommendations provided in their evaluations as well as examine their own potential cultural biases (Hicks, 2004).

Results from this study should be viewed as exploratory and descriptive rather than prescriptive in nature. As this was the first study examining quality in forensic evaluations in child welfare, results may not generalize. It is also possible that certain predictors that should have been included were not. This study only included data from New Jersey, and state child abuse laws may vary across the United States. Therefore, the results may not be representative of professional practice throughout the country. Furthermore, a lot of the studies related to quality of forensic evaluations at this present time have been conducted in countries other than the United States, such as the Netherlands and Portugal (da Silva Guerreiro et al., 2014; Duits et al., 2012). Despite the exploratory nature of the findings, some of the trends discovered in the current study's data provide valuable information and warrant further discussion.

Strengths of the Study

The current study found that certain factors within a forensic evaluation may predict elements of high-quality. The regression models revealed significant results, however, Nagelkerke r^2 , which indicates the power of the explanation of the model, was relatively weak in both models. Additionally, some of the odds ratios were relatively weak and not statistically

significant. Even with the weak power of the explanation of the model in both regression models, the results add to the literature in several ways. This study provides a glimpse of the possible relationships between elements of an evaluation and how they may contribute to measures of quality of the evaluation. It also is one of the first studies of its kind to collect data on forensic evaluations conducted in the field of child welfare. It provides insight into what kind of information is lacking in evaluations, such as specific demographic information like race, ethnicity, and language spoken by both the subject and the evaluator. Studies like the current study that provide this type of information are important because any addition to the literature can help promote positive outcomes for vulnerable children and families (Mallon & Hess, 2014).

The QI Tool used in this study can also help promote dialogue about quality of forensic evaluations both among psychologists and among all stakeholders. A standardized quality assessment tool creates a “language” in which to speak about these evaluations that did not exist before. Forensic evaluators can use tools such as this instrument to aid in self-assessment. In combination with externally mandated quality control, self-assessment procedures may be useful to forensic evaluators (Wettstein, 2005, 2010; Zwartz, 2018). Further research is needed to investigate whether the QI Tool can support forensic training, peer review, and supervision in forensic areas other than child welfare. However, it provides a strong foundation to do so within the area of child maltreatment evaluations.

Clinical Implications

There have been many identified areas for improvement and common errors in forensic evaluation (Grisso, 2010; Neal & Grisso, 2014; Wettstein 2005, 2010). Forensic evaluators should use studies like this, as well as the studies available describing quality improvement in forensic evaluation and the results of the current study to monitor their work. Witt (2010)

created a forensic report checklist, with the idea that checklists can reduce errors in a wide range of complex tasks. He emphasizes how checklists differ from templates, in that they also stipulate specific steps in the process that need to be completed, in addition to headers and topics. Witt (2010) utilized Grisso's (2010) article about typical errors in forensic evaluations to formulate a checklist encompassing the ten most common errors. Checklists could be useful to forensic examiners in child welfare, especially if they are adapted to speak to the specific areas that must be addressed in parenting and child abuse and neglect evaluations (Zwartz, 2018). Ethical and legal requirements of the assessment process can also be incorporated into this checklist (Fuger et al., 2014). According to results from the current study, aspects about culture and diagnosis should also be added to a checklist for completing forensic evaluations in the field of child welfare. However, as Neal and Grisso (2014) state, little is known about the degree to which structured tools to assist professional judgement in forensic evaluation are used.

As much as self-monitoring can lead to opportunities for higher quality forensic evaluations to be written, peer review, training, and feedback may be even more powerful (Neal & Brodsky, 2016). Feedback from the courts and other involved parties on the usability of forensic evaluations may be an advantageous process for all parties involved (Robinson & Acklin, 2010), and may provide more insight as to which aspects of a report may be predictors of measures of quality of the report (Lander & Heilbrun, 2009). Education and training of forensic evaluators should improve report quality, but research shows that this training must be ongoing (Herman, 2005; Robinson & Acklin, 2010). Studies that examine interrater reliability between different evaluators can prove to be especially useful in the process of understanding what is most important in regard to improving quality of forensic evaluations (Fuger et al., 2014; Nguyen et al., 2011). Interrater reliability among the decision reached in the evaluation is often

equated with quality, and it has been found that the most important factor that leads to an increase in interrater reliability is the training of forensic examiners (Acklin et al., 2015). Interestingly enough, studies have noted that examiner experience and confidence is not always correlated with interrater reliability in some types of forensic evaluations (Sutherland et al., 2012). The NJCC, the provider of the data for the current study, can also be used as an example of an institution that helps to foster ongoing training and peer review. As the NJCC reviews evaluations in different areas of NJ, it holds dissemination meetings to discuss the data and results and invites all evaluators who were reviewed using the QI Tool to attend these meetings.

Substantial improvement in measures of quality in forensic evaluations in child welfare is likely to require a systemic approach. General and specialty guidelines exist (APA 2013a, 2013b), however they may be a bit aspirational and non-specific, thus falling short in dictating what exactly indicates an evaluation is high-quality. The field of forensic evaluation in general lacks a standard of practice (Acklin et al., 2015), which is a critical issue that points to the need of rigorous studies examining areas for improvement such as the current study. The QI Tool used in this study and similar tools in studies conducted in other countries (da Silva Guerreiro et al., 2014; Duits et al., 2012) displays the need for quality assessment tools with good psychometric properties, which can then lead to empirically sound quality improvement analyses. There is strong evidence that examining individual principles and factors, similar to what the current study accomplished, in forensic evaluations leads to validation of which areas should be included in these evaluations (Lander & Heilbrun, 2009). However, research such as the current study must continue to be conducted in order to determine which factors of evaluations in child welfare cases are imperative to include to ensure best practices are being employed.

Ethical Implications

Forensic psychological evaluations in child welfare are more than a report written to influence the courts, as ethical and practice principles must be considered. Previous studies in forensic psychology and psychiatry have examined how reports in the United States are typically written (Budd, 2005; Grisso, 2010; Heilbrun et al., 2008; Melton et al., 2017; Young, 2016). Young (2016) conducted a survey of different approaches to forensic report writing and psychological ethical approaches. He summarized all ethical principles from prior research done by Heilbrun et al. (2008) that he believed apply to forensic reports. Young (2016) argues that forensic assessors need to use scientific reasoning as they judge the subject's clinical conditions and functional abilities because this is what the APA's forensic guidelines (2013b) and psychological ethics support. Dror and Murrie (2018) emphasize that forensic psychologists and psychiatrists must consider their reliability and their "biasability" in order to arrive at ethically sound conclusions. These arguments speak to the notion that forensic psychological evaluations should be a more standardized, quality-focused process, from choosing instruments to arriving at legally relevant opinions. Neal and Grisso (2014) also posit that using standardized tools to improve clinical decision making can reduce bias, increase interrater reliability, and increase validity.

An important ethical concern brought about by the current study is that culture was not found to be integrated into a large portion of the evaluations. As DCF's Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (2012) state that evaluators must interpret results within the context of the diverse cultural and communities of NJ, it does not seem like a majority of the evaluations were conducted with strict compliance to ethical and procedural guidelines. There are different methods of child rearing within different cultures, and this should

be described in forensic evaluations in relation to state and federal laws (Aggarwal, 2012; Glancy et al., 2015). Despite the common knowledge that cultural discrimination may have a strong influence on outcomes of psychological and psychiatric evaluations, culture seems to remain poorly understood in forensic evaluations (Aggarwal, 2012; Alegría et al., 2008).

It has been suggested that forensic evaluators try to remove all empathy from their evaluation process in order to be completely objective about a situation. However, Mulay et al. (2018) conceptualize this as an ethical issue that warrants further investigation and debate within forensic evaluations in general. Empathy is a core component of most clinical psychology training, and may especially be relevant in the field of child maltreatment. There is a delicate balance that must be attempted to be made between evaluators managing their emotional reactions and evaluators understanding their subject's perspective and how stigma and culture can intertwine within the context of the subject's experience. As much as facts and test results need to be considered in forensic evaluations in child welfare, it may be ethically valuable for evaluators to also acknowledge their internal reactions and experiences of empathy for their subjects, whether it be towards the victim of maltreatment or the perpetrator.

Implications for Future Research

This study has many implications for future research in quality improvement of forensic psychological evaluations and psychological evaluations in general. Often, quality of psychological evaluations is overlooked in the field, possibly due to the focus being on other measures of quality of services, such as outcome measures and client feedback (Williams, 2016). Nonetheless, this study has illustrated that there is a need for improvement in the quality of forensic psychological evaluations in child welfare and a lack of literature on this topic. There is also a need for further investigation into what elements of an evaluation contribute to its quality.

There have been a few studies that attempt to measure forensic psychological evaluation quality (Duits et al., 2012; Fuger et al., 2014; Grisso, 2010; Nguyen et al., 2011), but this is not a ubiquitous process. Certain studies, such as the current study, had very nuanced grids and measures of quality (da Silva Guerreiro et al., 2014; Duits et al., 2012), and the development of a more generally applicable quality assessment and improvement tool would be an interesting potential research avenue.

It may be beneficial to incorporate practicing psychologists more into quality improvement and assessment initiatives (da Silva Guerreiro et al., 2014; Duits et al., 2012), as the current study did through peer review. The use of a quality framework and stakeholder input about this framework may provide valuable insights as to what domains of a report are most important in regard to measures of quality (Wettstein, 2005, 2010). Currently, stakeholder perspectives of measures of quality in evaluations in child welfare, including psychologists, court personnel and lawyers, child welfare workers, service providers, and families, are not typically known. Stakeholder perspectives and opinions can also be compared, as these reports should be usable for all involved parties. It would be interesting to compare evaluations completed by practitioners involved in this study to evaluations completed by others, possibly in different states, who were not included in the study. The results of this study also suggest that increasing the amount of standardization and evaluation training in child welfare may increase evaluation quality, as there seemed to be more child evaluations attaining a higher score on measures of quality. It may be helpful for the QI Tool to investigate and measure quality improvement throughout various years, such as Duits and colleagues' (2012) study similarly examined. However, the literature on the abovementioned topics remains unclear and should be investigated further.

Few studies have addressed the impact of ethnic, cultural, and linguistic factors on the forensic practice of evaluations in child welfare (Duits et al., 2012; Kirmayer et al., 2007). Additionally, there was not enough data in this study's fairly large sample to analyze any relationships between ethnic, cultural, or linguistic diversity with forensic evaluation quality. As it is known that misdiagnosis is common among individuals from diverse backgrounds, it is especially important for future research to investigate culture and how it is integrated into forensic evaluations in child welfare (Liang et al., 2016). As more research is conducted, it will become clearer as to whether or not there are differences in diagnostic prevalence among different racial and ethnic groups as opposed to consequences of misdiagnosis. Forensic evaluators may not be able to control the interactions present between dominant and nondominant ethnic groups within the justice system, however they should be presenting enough information about an individual's culture that will allow for minimal exercise of stereotypes based on ethnicity and culture (Hicks, 2004). Exploration of whether or not this is occurring is particularly relevant to the field of child welfare, as the current study was not able to address this due to lack of information provided in the current sample's evaluations.

References

- Acklin, M. W., Fuger, K., & Gowensmith, W. (2015). Examiner agreement and judicial consensus in forensic mental health evaluations. *Journal of Forensic Psychology Practice, 15*(4), 318-343.
- Aggarwal, N. K. (2012). Adapting the cultural formation for clinical assessments in forensic psychiatry. *Journal of the American Academy of Psychiatry and the Law.*
- Alegria, M., Nakash, O., Lapatin, S., Oddo, V., Gao, S., Lin, J., & Normand, S. L. (2008). How missing information in diagnosis can lead to disparities in the clinical encounter. *Journal of Public Health Management and Practice: JPHMP, 14*(Suppl), S26.
- Alegria, M., Vallas, M., & Pumariega, A. J. (2010). Racial and ethnic disparities in pediatric mental health. *Child and Adolescent Psychiatric Clinics, 19*(4), 759-774.
- Allan, A., & Grisso, T. (2014). Ethical principles and the communication of forensic mental health assessments. *Ethics & Behavior, 24*(6), 467-477.
- American Psychiatric Association. (2013a). Cautionary statement for forensic use of DSM-5. In *Diagnostic and statistical manual of mental disorders* (5th ed.). doi:10.1176/appi.books.9780890425596.744053
- American Psychiatric Association. (2013b). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>
- American Psychological Association. (2013a). Guidelines for psychological evaluations in child protection matters. *The American Psychologist, 68*(1), 20.
- American Psychological Association. (2013b). Specialty guidelines for forensic psychology. *The American Psychologist, (68) 1*, 377-402.
- Archer, R. P., Wheeler, E. M., & Vauter, R. A. (2016). Empirically supported forensic

- assessment. *Clinical Psychology: Science and Practice*, 23(4), 348-364.
- Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, 19(1), 3-11.
- Baum, F., MacDougall, C., & Smith, D. (2006). Participatory action research. *Journal of Epidemiology & Community Health*, 60(10), 854-857.
- Berliner, L., Fitzgerald, M. M., Dorsey, S., Chaffin, M., Ondersma, S. J., & Wilson, C. (2015). Report of the APSAC Task Force on evidence-based service planning guidelines for child welfare. *Child Maltreatment*, 20(1), 6-16.
- Boehnlein, J. K., Schaefer, M. N., & Bloom, J. D. (2005). Cultural considerations in the criminal law: the sentencing process. *Journal of the American Academy of Psychiatry and the Law Online*, 33(3), 335-341.
- Bonin, L. (2018). Quality improvement in health care: the role of psychologists and psychology. *Journal of Clinical Psychology in Medical Settings*, 25(3), 278-294.
- Buchanan, A., & Norko, M. (2013). The forensic evaluation and report: an agenda for research. *The Journal of the American Academy of Psychiatry and the Law*, 41(3), 359-365.
- Budd, K. S. (2005). Assessing parenting capacity in a child welfare context. *Children and Youth Services Review*, 27(4), 429-444.
- Centers for Disease Control and Prevention. (2014). *Understanding child maltreatment*. Retrieved from <https://www.cdc.gov/violenceprevention/pdf/understanding-cm-factsheet.pdf>

- Cheng, H. G., & Phillips, M. R. (2014). Secondary analysis of existing data: opportunities and implementation. *Shanghai Archives of Psychiatry*, 26(6), 371–375.
<http://doi.org/10.11919/j.issn.1002-0829.214171>
- Child Welfare Information Gateway. (2018.) Child maltreatment 2016: Summary of key findings. Washington, DC: U.S. Department of Health and Human Services, Children's Bureau. <https://www.acf.hhs.gov/cb/resource/child-maltreatment-2016>
- Chiu, E. Y. (2014). Psychological testing in child custody evaluations with ethnically diverse families: Ethical concerns and practice recommendations. *Journal of Child Custody*, 11(2), 107-127.
- Combalbert, N., Andronikof, A., Armand, M., Robin, C., & Bazex, H. (2014). Forensic mental health assessment in France: Recommendations for quality improvement. *International Journal of Law and Psychiatry*, 37(6), 628-634.
- Conroy, M. A. (2006). Report writing and testimony. *Applied Psychology in Criminal Justice*, 2(3), 237-260.
- Coohey, C., & Braun, N. (1997). Toward an integrated framework for understanding child physical abuse. *Child Abuse & Neglect: The International Journal*, 21(11), 1081–1094.
[https://doi.org/10.1016/S0145-2134\(97\)00067-7](https://doi.org/10.1016/S0145-2134(97)00067-7)
- Cradock, G. (2011). Thinking goudge: Fatal child abuse and the problem of uncertainty. *Current Sociology*, 59(3), 362-378.
- Craddock, N., & Mynors-Wallis, L. (2014). Psychiatric diagnosis: impersonal, imperfect and important. *The British Journal of Psychiatry*, 204(2), 93-95.
- Darling, N., & Steinberg, L. (1993). Parenting style as context: An integrative model. *Psychological Bulletin*, 113(3), 487.

- da Silva Guerreiro, J., Casoni, D., & Costa Santos, J. (2014). Relevance and coherence as measures of quality in forensic psychological reports. *Psychiatry, Psychology and Law*, 21(6), 890-902.
- Davis, K. M., & Lister, M. B. (2019). Conducting disability evaluations with a forensic perspective: The application of criminal responsibility evaluation guidelines. *Psychological Injury and Law*, 12(1), 52-63.
- deMause, L. (1988). *The History of Childhood: The Untold History of Child Abuse*. New York: Peter Bedrick.
- Detlaff, A. J., Rivaux, S. L., Baumann, D. J., Fluke, J. D., Rycraft, J. R., & James, J. (2011). Disentangling substantiation: The influence of race, income, and risk on the substantiation decision in child welfare. *Children and Youth Services Review*, 33(9), 1630-1637.
- Dror, I. E., & Murrie, D. C. (2018). A hierarchy of expert performance applied to forensic psychological assessments. *Psychology, Public Policy, and Law*, 24(1), 11-23.
- Dubowitz, H. (2014). Child Neglect. *Pediatric Annals*, 43(11), 444-445.
<https://doi.org/10.3928/00904481-20141022-07>
- Dubowitz, H. (1999). *Neglected Children : Research, Practice, and Policy*. Thousand Oaks, Calif: Sage Publications.
- Duits, N., van der Hoorn, S., Wiznitzer, M., Wettstein, R. M., & de Beurs, E. (2012). Quality improvement of forensic mental health evaluations and reports of youth in the Netherlands. *International Journal of Law and Psychiatry*, 35(5-6), 440-444.
- Etter, D., & Rickert, V. (2013). The complex etiology and lasting consequences of child

- maltreatment. *Journal of Adolescent Health*, 53(4), S39–S41.
<https://doi.org/10.1016/j.jadohealth.2013.07.010>
- Finkelhor, D., & Araji, S. (1986). Explanations of pedophilia: A four factor model. *Journal of Sex Research*, 22(2), 145–161. <https://doi.org/10.1080/00224498609551297>
- Fuger, K. D., Acklin, M. W., Nguyen, A. H., Ignacio, L. A., & Gowensmith, W. N. (2014). Quality of criminal responsibility reports submitted to the Hawaii judiciary. *International Journal of Law and Psychiatry*, 37(3), 272-280.
- Forsythe, M. (2018). [Payment detail sheet FY2018]. Unpublished raw data.
- Glancy, G. D., Ash, P., Bath, E. P., Buchanan, A., Fedoroff, P., Frierson, R. L., ... & Norko, M. (2015). AAPL Practice Guideline for the Forensic Assessment. *The Journal of the American Academy of Psychiatry and the Law*, 43(2 Suppl), S3.
- Gold, L. H. (2013) Mental health disability: A model for assessment. In *Clinical guide to mental disability evaluations* (pp. 3-35). Springer, New York, NY.
- Goodman-Delahunty, J., & Dhami, M. K. (2013). A forensic examination of court reports. *Australian Psychologist*, 48(1), 32-40.
- Gowensmith, W. N., Sledd, M., & Sessarego, S. (2015). The impact of stringent certification standards on forensic evaluator reliability: Further analysis. In *annual meeting of the American Psychology-Law Society, San Diego, CA*.
- Gowensmith, W. N., Sessarego, S. N., McKee, M. K., Horkott, S., MacLean, N., & McCallum, K. E. (2017). Diagnostic field reliability in forensic mental health evaluations. *Psychological Assessment*, 29(6), 692.
- Greenberg, S., & Shuman, D. (2007). When worlds collide: Therapeutic and forensic roles. *Professional Psychology: Research and Practice*, 38(2), 129-132.

- Greenland, S., Mansournia, M. A., & Altman, D. G. (2016). Sparse data bias: a problem hiding in plain sight. *BMJ*, 352, i1981.
- Grisso, T. (1987). The economic and scientific future of forensic psychological assessment. *American Psychologist*, 42(9), 831.
- Grisso, T. (2010). Guidance for improving forensic reports: A review of common errors. *Open Access Journal of Forensic Psychology*, 2, 102–115.
- Hacking, I. (1991). The making and molding of child abuse. *Critical Inquiry*, 17(2), 253-288.
- Harvey, L. (2007). Epistemology of quality. *Perspectives in Education*, 25(3), 1-13.
- Hearn, J. (2011). Unmet needs in addressing child neglect: Should we go back to the drawing board? *Children and Youth Services Review*, 33(5), 715-722.
- Heilbrun, K. (2001). *Principles of Forensic Mental Health Assessment* (1st ed. 2001.). <https://doi.org/10.1007/b100329>
- Heilbrun, K. (2003). Principles of forensic mental health assessment: implications for the forensic assessment of sexual offenders. *Annals of the New York Academy of Sciences*, 989, 167–184.
- Heilbrun, K., DeMatteo, D., & Marczyk, G. (2004). Pragmatic psychology, forensic mental health assessment, and the case of Thomas Johnson: Applying principles to promote quality. *Psychology, Public Policy, and Law*, 10(1-2), 31.
- Heilbrun, K., DeMatteo, D., Marczyk, G., & Goldstein, A. M. (2008). Standards of practice and care in forensic mental health assessment: Legal, professional, and principles-based consideration. *Psychology, Public Policy, and Law*, 14(1), 1.
- Heilbrun, K., Grisso, T., & Goldstein, A. M. (2009). *Foundations of forensic mental health assessment*. New York: Oxford University Press.

- Herman, S. (2005). Improving decision making in forensic child sexual abuse evaluations. *Law and Human Behavior*, 29(1), 87.
- Hicks, J. W. (2004). Ethnicity, race, and forensic psychiatry: are we color-blind? *Journal of the American Academy of Psychiatry and the Law Online*, 32(1), 21-33.
- Humphrey, C. (2015). Evil, child abuse and the caring professions. *Journal of Religion and Health*, 54(5), 1660-1671.
- Jackson, R. H. (1995). The history of childhood accident and injury prevention in England: background to the foundation of the Child Accident Prevention Trust. *Injury Prevention*, 1(1), 4.
- Jud, A., Fegert, J. M., & Finkelhor, D. (2016). On the incidence and prevalence of child maltreatment: a research agenda. *Child and Adolescent Psychiatry and Mental Health*, 10(1), 17.
- Kempe, C. H., Silverman, F. N., Steele, B. F., Droegemueller, W., & Silver, H. K. (1985). The battered-child syndrome. *Child Abuse & Neglect*, 9(2), 143-154.
- Kim, B., & Maguire-Jack, K. (2015). Community interaction and child maltreatment. *Child Abuse & Neglect*, 41, 146-157.
- Kirmayer, L. J., Rousseau, C., & Lashley, M. (2007). The place of culture in forensic psychiatry. *Journal of American Academy of Psychiatry and the Law*, 35(1), 98.
- Kuehnle, K., Coulter, M., & Firestone, G. (2000). Child protection evaluations: The forensic stepchild. *Family Court Review*, 38(3), 368-391.
- Lander, T. D., & Heilbrun, K. (2009). The content and quality of forensic mental health assessment: Validation of a principles-based approach. *International Journal of Forensic Mental Health*, 8(2), 115-121.

- Lansford, J. E., Godwin, J., Tirado, L. M. U., Zelli, A., Al-Hassan, S. M., Bacchini, D., ... & Di Giunta, L. (2015). Individual, family, and culture level contributions to child physical abuse and neglect: A longitudinal study in nine countries. *Development and Psychopathology*, 27(4pt2), 1417-1428.
- Leventhal, J., & Edwards, G. (2017). Flawed theories to explain child physical abuse: What are the medical-legal consequences? *JAMA*, 318(14), 1317–1318.
<https://doi.org/10.1001/jama.2017.11703>
- Liang, J., Matheson, B. E., & Douglas, J. M. (2016). Mental health diagnostic considerations in racial/ethnic minority youth. *Journal of Child and Family Studies*, 25(6), 1926-1940.
- Maguire-Jack, K., & Font, S. (2017). Community and individual risk factors for physical child abuse and child neglect: Variations by poverty status. *Child Maltreatment*, 22(3), 215–226. <https://doi.org/10.1177/1077559517711806>
- Mallon, G. P., & Hess, P. M. (2014). *Child welfare for the twenty-first century: A handbook of practices, policies, & programs*. Columbia University Press.
- McCallum, K. E., & Gowensmith, W. N. (2019). Tipping the scales of justice: the role of forensic evaluations in the criminalization of mental illness. *CNS Spectrums*, 1-7.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30.
- Melton, G. B., Petrila, J., Poythress, N. G., Slobogin, C., Otto, R. K., Mossman, D., & Condie, L. O. (2017). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers*. Guilford Publications.
- Mulay, A. L., Mivshek, M., Kaufman, H., & Waugh, M. H. (2018). The ethics of empathy:

- Walking a fine line in forensic evaluations. *Journal of Forensic Psychology Research and Practice*, 18(4), 320-336.
- Munro, B. H. (2005). *Statistical methods for health care research* (Vol. 1). Lippincott, Williams & Wilkins.
- Neal, T., & Brodsky, S. L. (2016). Forensic psychologists' perceptions of bias and potential correction strategies in forensic mental health evaluations. *Psychology, Public Policy, and Law*, 22(1), 58.
- Neal, T. M., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior*, 41(12), 1406-1421.
- New Jersey Coordination Center [NJCC] (2018). Quality Improvement Tool Codebook. New Brunswick, NJ: Author.
- New Jersey Department of Children and Families. (2012). Guidelines for Expert Evaluations in Child Abuse/Neglect Proceedings (Mental Health). Trenton, NJ.
- Nguyen, A. H., Acklin, M. W., Fuger, K., Gowensmith, W. N., & Ignacio, L. A. (2011). Freedom in paradise: Quality of conditional release reports submitted to the Hawaii judiciary. *International Journal of Law and Psychiatry*, 34(5), 341-348.
- Nicholson, R. A., & Norwood, S. (2000). The quality of forensic psychological assessments, reports, and testimony: Acknowledging the gap between promise and practice. *Law and Human Behavior*, 24(1), 9-44.
- Parton, N. (1979). The natural history of child abuse: a study in social problem definition. *The British Journal of Social Work*, 9(4), 431-451.
- Pence, D. (2011). Child abuse and neglect investigation. In J. E. B. Myers (Ed.), *The APSAC handbook on child maltreatment* (pp. 325– 336). Thousand Oaks, CA: Sage.
- Pickar, D. B., & Kaufman, R. L. (2013). The child custody evaluation report: Toward an

- integrated model of practice. *Journal of Child Custody*, 10(1), 17-53.
- Raghavan, R., & Alexandrova, A. (2015). Toward a theory of child well-being. *Social Indicators Research*, 121(3), 887-902.
- Robinson, R., & Acklin, M. W. (2010). Fitness in paradise: Quality of forensic reports submitted to the Hawaii judiciary. *International Journal of Law and Psychiatry*, 33(3), 131-137.
- Rutgers School of Social Work Institute for Families. (n.d.). New Jersey Coordination Center for Child Abuse and Neglect Forensic Evaluation and Treatment. Retrieved from <https://socialwork.rutgers.edu/centers/institute-families/new-jersey-coordination-center-child-abuse-and-neglect-forensic-evaluation-and-treatment>
- Sadoff, R. L., & Dattilio, F. M. (2012). Formal training in forensic mental health: Psychiatry and psychology. *International Journal of Law and Psychiatry*, 35(5-6), 343-347.
- Sanschagrín, K. A. (2006). The quality of forensic mental health assessments of juvenile offenders: An empirical investigation.(ProQuest Information & Learning). *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 66(11), 6292.
- Scannapieco, M., & Connell-Carrick, K.(2002). Focus on the first years: An eco-developmental assessment of child neglect for children 0 to 3 years of age. *Children and Youth Services Review*, 24(8), 601–621. [https://doi.org/10.1016/S0190-7409\(02\)00210-4](https://doi.org/10.1016/S0190-7409(02)00210-4)
- Schlomer, G. L., Bauman, S., & Card, N. A. (2010). Best practices for missing data management in counseling psychology. *Journal of Counseling Psychology*, 57(1), 1.
- Sullivan, G. M. (2011). A primer on the validity of assessment instruments. *Journal of Graduate Medical Education*, 3(2), 119-120.
- Svevo-Cianci, K., & Lee, Y. (2010). Twenty years of the Convention on the Rights of the Child: Achievements in and challenges for child protection implementation, measurement and

- evaluation around the world. *Child Abuse and Neglect*, 34(1), 1-4.
- Sykes, J. (2011). Negotiating stigma: Understanding mothers' responses to accusations of child neglect. *Children and Youth Services Review*, 33(3), 448-456.
<https://doi.org/10.1016/j.childyouth.2010.06.015>
- Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(3), 227.
- Tzeng, O., Jackson, J., & Karlson, H. (1991). *Theories of child abuse and neglect: Differential perspectives, summaries, and evaluations*. New York: Praeger.
- U.S. Department of Health and Human Services. (2015). Child maltreatment 2014. Washington, DC: Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau.
- Wettstein, R. M. (2005). Quality and quality improvement in forensic mental health evaluations. *Journal of the American Academy of Psychiatry and the Law Online*, 33(2), 158-175.
- Wettstein, R. M. (2010). Commentary: Conceptualizing the forensic psychiatry report. *Journal of the American Academy of Psychiatry and the Law Online*, 38(1), 46-48.
- Williams, G., Tonmyr, L., Jack, S. M., Fallon, B., & MacMillan, H. L. (2011). Determinants of maltreatment substantiation in a sample of infants involved with the child welfare system. *Children and Youth Services Review*, 33(8), 1345-1353.
- Williams, J. (2016). Quality assurance and quality enhancement: Is there a relationship? *Quality in Higher Education*, 22(2), 97-102.

- Wills, C. D., & Gold, L. H. (2014). Introduction to the special section on DSM-5 and forensic psychiatry. *Journal of the American Academy of Psychiatry and the Law Online*, 42(2), 132-135.
- Witt, P. H. (2010). Forensic report checklist. *Open Access Journal of Forensic Psychology*, 2, 233-240.
- Wortzel, H. S. (2013). The DSM-5 and forensic psychiatry. *Journal of Psychiatric Practice*, 19(3), 238-241.
- Young, G. (2016). Psychiatric/psychological forensic report writing. *International Journal of Law and Psychiatry*, 49, 214-220.
- Zapf, P. A., & Dror, I. E. (2017). Understanding and mitigating bias in forensic evaluation: Lessons from forensic science. *International Journal of Forensic Mental Health*, 16(3), 227-238.
- Zumbach, J., Wetzels, P., & Koglin, U. (2018). Predictors of psychological recommendations in child protection evaluation. *Child Abuse & Neglect*, 84, 196-204.
- Zwartz, M. (2018). Report writing in the forensic context: Recurring problems and the use of a checklist to address them. *Psychiatry, Psychology and Law*, 1-11.

Appendix A

New Jersey Department of Children and Families' Guidelines for Expert Evaluations in Child
Abuse/Neglect Proceedings (Mental Health)

New Jersey Department of Children and Families

**Guidelines for Expert Evaluations in
Child Abuse/Neglect Proceedings (Mental Health)**

November 8, 2012

Commissioner Allison Blake, PhD LSW

**Child Abuse and Neglect Mental Health Evaluation and Treatment Advisory Group
Member List**

Chair: Debra Lancaster

Director
Office of Child and Family Health
NJ Department of Children and Families

Karen Baldoni

Director
Office of Contract Administration
NJ Department of Children and Families

Mary Beirne, MS, EdD, MD

Chief Child/Adolescent Psychiatrist
NJ Department of Children and Families

Lauren Carlton

Assistant Attorney General
NJ Division of Law

Nancy Carre-Lee

Assistant Director
Division of Child Protection & Permanency
NJ Department of Children and Families

Anthony V. D'Urso, Psy.D.

Section Chief and Supervising Psychologist
Audrey Hepburn Children's House
Associate Professor - Graduate Clinical Psychology
Montclair State University

John Esmerado

Assistant Prosecutor
Union County Child Advocacy Center

Amanda Farley

Program Support Specialist
Office of Child and Family Health
NJ Department of Children and Families

Jeffrey Guenzel

Deputy Commissioner
NJ Department of Children and Families

Kristin Huddy

Program Support Specialist
Office of Contract Administration
NJ Department of Children and Families

Adrienne Jackson, MSW

Executive Coordinator
NJ Task Force on Child Abuse & Neglect

Walter Kalman, MSW, LSW

Executive Director
National Association of Social Workers, NJ Chapter

Barry A. Katz, Ph.D.

NJ Psychological Association

Anthony P. Kearns, III

Prosecutor
Hunterdon County Prosecutor's Office

Julie Lippman, Psy.D.

Evaluation Services Director
CARES Institute

Keri Logosso-Misurell, Esq.

Board Member
New Jersey Children's Alliance

Joseph Ribsam Jr., Esq.

Assistant Commissioner
NJ Department of Children and Families

Brian Ross, Esq.

Assistant Director
Office of Policy and Regulatory Development
NJ Department of Children and Families

Aileen Torres, Ph.D.

Past-President
Latino Psychological Association of New Jersey

Kara Wood

Director
Division of Child Protection & Permanency
NJ Department of Children and Families

Table of Contents

I. Introduction	4
II. Definition/Application	4
III. General Principles and Guidelines	5
1. The Role and Function of Forensic Evaluation in Child Welfare Matters	5
2. General Competencies of Expert Evaluators	6
3. Procedural Guidelines: Conducting an Evaluation	7
IV. Best Practices for Expert Forensic Evaluations	9
1. During an Investigation	9
2. During Permanency Planning/Hearings	11
3. During Litigation of Guardianship Complaints	14

I. Introduction

This is the Department's first comprehensive effort to address the use of expert evaluations in child welfare and child protective services proceedings. These guidelines lay out best practices for forensic evaluations and assessments that may be needed during child welfare investigations, to assist with permanency planning, or during litigation of guardianship complaints.

Child abuse and neglect cases are often complex. Expert consultants are frequently used to assist caseworkers, attorneys, law guardians, judges, and parents in making determinations, case planning, and decision making. The experts' services are often in the form of forensic evaluations of the mental health status and/or capabilities of the parents of dependent children. In addition, an evaluator may assess a child's behavioral functioning or developmental status as well. CP&P and the courts often rely on these evaluations and recommendations for effective case planning and to guide the court's decision making process.

In developing the guidelines that follow, the Department reviewed and analyzed professional guidelines and the work of other states, and convened an interdisciplinary group of experts to form DCF's Advisory Group on Child Abuse and Neglect Mental Health Evaluation and Treatment.

The role of the Advisory Group was to assist in formulating a framework that is flexible enough to accommodate differences in disciplines while providing clear practice guidelines that address the questions to be asked, the information required, the tools necessary to inform the evaluation, the credentials and qualifications of the evaluator, and the essential components of the evaluation itself.

The guidelines that follow are intended to improve the quality of expert forensic evaluations provided for CP&P and the courts, as well as the ability of stakeholders involved in child welfare proceedings and child protective service matters to make better use of them. It is clear that representatives of different disciplines with differing philosophical orientations will have varying approaches to the task of providing a forensic assessment. Each unique discipline will organize their work in a way that reflects their individual expertise. These guidelines are not meant to supplant the professional judgment of evaluators regarding their response to the unique features of each case.

The first sections of this document are general guidelines, followed by more specific recommended practices.

II. Definition/Application

For the purpose of these guidelines, a forensic evaluation in child welfare proceedings and child protective service matters is an evaluation necessary to assist the court and/or CP&P in case planning, or to resolve a case. A forensic evaluation may be requested by CP&P, by another party to a proceeding, or the court. Any evaluation that may reasonably be expected to be submitted to the court is termed forensic. Although forensic evaluations may contain treatment

recommendations, the primary function of the forensic evaluation is to inform the parties and to assist the court in rendering decisions in child welfare cases.

These guidelines do not cover evaluations or assessments obtained primarily for mental health treatment purposes, substance abuse, anger management, psycho-sexual evaluation, or domestic violence, although any or all of these issues may be addressed in a forensic evaluation.

These guidelines recognize that, in child welfare cases, the emphasis is on the safety, permanency, and well-being of the child.

III. General Principles and Guidelines¹

1. The Role and Function of Forensic Evaluations in Child Welfare Matters

The primary function of an evaluation is to provide a report that contains relevant, professionally sound observations, results and opinions in matters where a child's health and welfare may have been harmed or placed at risk of harm. To ensure the reliability of the evaluator's conclusions all opinions that are rendered must be given within a reasonable degree of medical/psychological/clinical certainty. The specific purposes of the evaluation generally will be determined by the referral questions and/or concerns provided to the evaluator by the referring party or parties. When the child already has been found by the court to be at risk of harm, the evaluation of the parent(s) generally identifies interventions intended to reduce future risk to the child, and often focuses on rehabilitation recommendations designed to protect the child and help the family. An additional purpose of such an evaluation may be to make recommendations for interventions that promote the psychological and physical well-being of the child, and, when appropriate, facilitate the safe reunification of the child with the parent. Consistent with State law, evaluators appreciate the value of expediting family reunification, when possible and safe, while they also understand the value of other permanent plans when reunification is not possible.

The evaluation addresses the particular psychological, behavioral, and developmental needs of the child and/or parent(s). Relevant issues may include, but are not limited to, abuse or neglect of the child, safety, parental capabilities, or reunification or other permanency plans. In considering psychological factors affecting the health and welfare of the child, evaluators may focus on caregiver capacities in the context of the psychological and developmental needs of the child. This may involve an assessment of:

- The adult's capacities for parenting, including those attributes, skills, strengths and abilities most relevant to abuse and/or neglect concerns;

¹ Washington State Supreme Court Commission on Children in Foster Care. "Guidelines for Expert Evaluations in Child Welfare Proceedings." Washington, 2007. Available online at, [http://www.courts.wa.gov/committee/pdf/Guidelines%20for%20Expert%20Evaluations%20in%20Child%20Welfare%20Proceedings%20\(2007\).pdf](http://www.courts.wa.gov/committee/pdf/Guidelines%20for%20Expert%20Evaluations%20in%20Child%20Welfare%20Proceedings%20(2007).pdf)

- The psychological functioning, behavioral, and developmental needs of the child, particularly with regard to vulnerabilities and special needs of the child, as well as the quality of the child's attachment to the parent(s) and the possible developmental and emotional effects of separation from the parent(s), siblings, extended family members, and other caregivers;
- The current and potential functional abilities of the parent(s) and, when necessary for resolution of the case, other relatives, to meet the needs of the child; and/or
- The need for and likelihood of success of clinical or other interventions for identified problems, which may include recommendations regarding treatment modalities and objectives, frequency of services, specialized interventions, parent education, and the child's placement.

2. General Competencies of Expert Evaluators

Evaluators should gain and maintain specialized competence. Expert evaluators in child protection matters are aware that special competencies and knowledge are necessary for the undertaking of such evaluations. Competence in performing expert evaluations of children, adults and families is necessary but not sufficient. Education, training, experience and/or supervision in the areas of forensic practice, child and family development, child and adult psychopathology, the impact of separation on the child, the nature and consequences of different types of child abuse and neglect, and the significance of human differences may help to prepare evaluators to participate competently in expert evaluations in child protection matters.

Evaluators:

- Use current knowledge of scholarly and professional developments, consistent with generally accepted clinical and scientific practice, in selecting evaluation methods and procedures² and are aware of evidence-based practices.
- Strive to become familiar with applicable legal and regulatory standards and procedures, including local State and Federal laws governing child protection issues. These may include laws and regulations addressing child abuse, neglect, and termination of parental rights.
- Describe the scientific basis for their judgments or recommendations, and state when their judgments or recommendations may expand on, or not be fully supported by, currently accepted clinical and scientific practice.
- Are aware of, and develop their knowledge and special competencies for, evaluation of specific populations including, but not limited to, issues related to literacy, the needs of persons who do not speak English, sensory impairment, psychological disorders, and developmental impairments.
- Should be fluent in the child's/parent's native language, when possible (have experience using a court appointed interpreter, if language presents a difficulty).

² Note: Examples of standard setting organizations include American Psychological Association, the National Association of Family and Conciliation Courts, The American Academy of Child and Adolescent Psychiatry and others.

- Have appropriate qualifications to conduct an evaluation and/or to testify at court, including language, cultural competency, and other qualifications specified in CP&P contracts.
- Should be competent in the cultural norms of the child/parent being evaluated.
- Utilize language and culturally correct testing.
- Have expertise in working with relevant clinical populations, including:
 - Children;
 - Sex offenders;
 - Domestic violence victims and batterers;
 - Persons with developmental disabilities; and,
 - Persons with psychiatric/neurological/neuropsychiatric diagnosis.
- Have expertise with the instruments employed, including psychological and intellectual tests that will need to be interpreted by a licensed psychologist, who is familiar with the norms and the uses of that test with the relevant population.
- Are experts in the use of appropriate interview techniques.
- Must not serve as an expert evaluator if they are the treating professional.

Evaluators must be aware of personal and societal biases and engage in nondiscriminatory practice. Evaluators engaging in expert evaluations in child protection matters consider how biases regarding age, gender, race, ethnicity, national origin, religion, sexual orientation, disability, language, culture, and socioeconomic status may interfere with an objective evaluation and recommendations. Evaluators should be aware of the potential for defensiveness on the part of participants, given the circumstances, and must take this into account when conducting the evaluation and upon making recommendations. Evaluators recognize and strive to overcome any such biases. If unable to overcome his or her own biases, the evaluator will either withdraw from the evaluation or seek assistance in completing the evaluation. When interpreting evaluation results, evaluators must be aware that there are diverse cultural and community methods of child rearing, and consider these in the context of the existing local State and Federal laws. Also, evaluators should use, whenever available, tests validated with populations similar to those being evaluated.

Evaluators avoid multiple relationships to maintain objectivity. In conducting expert evaluations in child protective matters, evaluators avoid multiple role relationships. Evaluators generally do not conduct forensic evaluations in child protection matters in which they have provided clinical services for the child or the immediate family, or have had other involvement that may compromise their objectivity. Providing clinical services to the child or other participants following an expert evaluation is discouraged. A treating professional can be called to testify, but should NOT recommend a permanency plan.

3. Procedural Guidelines: Conducting an Evaluation

Evaluators and referring parties understand that forensic evaluations in child welfare and child protection matters may present a wide variety of legal and/or ethical considerations. Evaluators and all parties appreciate the need for timeliness in child protection matters (e.g., response to evaluation referral, scheduling appointments, completion of reports).

The purpose of the evaluation should be clear upon referral and should outline the specific questions to be addressed by the evaluation. In all cases, the referring party or parties should clearly state the purpose of the evaluation in writing and pose specific questions to be addressed in the evaluation.

Based on the nature of the referral issues and questions, the scope of the evaluation is determined in the referral or by the court, in consultation with the evaluator. Once the referral questions and scope of the evaluation have been accepted by the evaluator, the expert evaluator chooses appropriate methods with which to address the questions. Evaluators may also identify relevant issues not anticipated in the referral questions that could enlarge the scope of the evaluation; these should be conveyed to the parties as early as possible. For issues outside the scope of the evaluator's competency, the evaluator considers recommending additional services or evaluations.

Evaluators inform participants about the disclosure of information and the limits of confidentiality.

- When an evaluation is court ordered, it is not privileged and the evaluator informs the individuals of the nature of the evaluation and that the evaluation will be distributed to other parties as provided by court order. Evaluators conducting an evaluation in child protection matters ensure that the participants, including the child (to the extent feasible), are aware of the limits of confidentiality for the evaluation results. If the public agency or court is paying for the evaluation, the evaluator so informs the individual.
- When an evaluation is not court ordered, evaluators performing evaluations in child protection matters should obtain informed consent from all adult participants, and children and youth consistent with their developmental capacity to understand.
- When an evaluation is obtained by a party in an abuse/neglect or termination proceeding without the apparent knowledge or consent of the child welfare agency, guardian ad litem, and/or the court, the evaluator should advise the party being evaluated of the need to obtain and review appropriate and relevant information from the child welfare agency, guardian ad litem, and/or the court.

Evaluators use multiple methods of data gathering. Evaluators generally use multiple methods of data gathering, including, but not limited to, clinical interviews, observation, and/or psychological testing that are sufficient to provide appropriate substantiation for their findings. Evaluators should review relevant reports (e.g., information from child protection agencies, social service providers, law enforcement agencies, health care providers, child care providers, schools, and institutions). In evaluating parental capacity to care for a particular child or the quality of the parent-child interaction, evaluators should make reasonable efforts to perform formal observations of the child together with the parent, unless such observation is not necessary to respond to the questions posed in the evaluation or to support the recommendations and conclusions of the evaluator. Evaluators in some circumstances may rely on formal observations conducted by other neutral and competent professionals. It is recognized that in some circumstances, parent-child observations may not be necessary. Also, in some

circumstances, it may not be advisable to require parent-child contact for purposes of the evaluation. For example, in cases where the safety or well-being of the child is clearly in jeopardy or parental contact with the child has been prohibited by the court. In such cases, the evaluator should note explicitly the reason(s) that a parent-child observation was not included. Evaluators may also interview extended family members and other individuals, when appropriate (e.g., caregivers, grandparents, and teachers). However, these should not be considered as substitutes for formal observation.

Evaluators are able to provide clarification and answer questions relating to the evaluation(s) completed. Once an evaluation is completed, the evaluator must be available to speak with CP&P staff such as the assigned caseworker if there are any questions or concerns regarding the evaluation.

IV. Best Practices for Expert Forensic Evaluations

Forensic evaluations may be needed at any point in time during the lifespan of a child protective services case. The need for a forensic evaluation may emerge during the course of an investigation to assist with developing understanding or seeking clarity around the allegations of child abuse/neglect. More commonly, mental health evaluations may be required to contribute to the decisions by the court of the Division made about placement, reunification, permanency, and visitation. Finally, forensic evaluations are typically required for guardianship (termination of parental rights) litigation.

1. During an Investigation

The Role and Function of Forensic Evaluations during an Investigation

During an investigation, evaluations may be needed to assist CP&P and the Courts in assessing whether abuse and/or neglect occurred. These evaluations are meant to assist in clarifying or gathering additional information for investigative purposes with the lens of an expert. When sufficient evidence or clarity about the case has been achieved through the investigative work of the CP&P caseworker via interviews and collateral review, or teamed efforts with law enforcement or others involved in the investigative process, it is often not necessary to engage the services of an expert for an evaluation during an investigation.

Forensic evaluations during the investigatory phase of the case may be warranted as part of the investigative efforts conducted by CP&P (and law enforcement). These situations most often include allegations of sexual abuse and emotional abuse/neglect. In addition, an evaluation during the initial involvement with a child may assist CP&P in determining the impact of an event on a child's psychological functioning.

Evaluations that may be required during the course of an investigation are almost always time sensitive matters. Thus, it is recommended that referrals be made as close to the point in time of the allegation or the occurrence of the alleged incident as possible:

- Evaluators should receive referrals within 10 working days of the report.
- An appointment by the evaluator should be granted within 10 working days of the referral.
- CP&P shall provide available background materials by the time of the evaluation.
- Evaluators should complete their reports and provide them to CP&P within 10 working days following completion of the evaluation.

These guidelines recommend that no more than 45 days pass between the initial referral to a provider for an assessment, to the date the written report, with recommendations, is provided to CP&P for review.

The Forensic Evaluation Process during an Investigation

In consultation with supervisory staff, and the DAG if litigation is contemplated or a complaint has been filed, CP&P caseworkers should select a provider who has the appropriate credentials to perform the evaluation. In many cases, child protection staff should access their Regional Diagnostic Treatment Center to conduct these evaluations. CP&P requires licensed individuals to conduct evaluations. In most cases, these will be licensed psychologists. When the impact of physiological factors, medical illness, medication, neurological, or psychiatric disorder is complex, an evaluation by a psychiatrist or physician may be necessary. In limited circumstances, an assessment by a LCSW may be appropriate.

The purpose of the evaluation during the investigatory phase of a case must be clear and should outline the specific questions to be addressed by the evaluator. Confirm with the evaluator the purpose of the evaluation. It is particularly important to limit the number of interviews or evaluations a child experiences for both validity reasons and to avoid re-traumatizing a child.

Investigation Evaluation Referral Questions:

- Is this child's presentation consistent with the allegation?
- To what degree has the child been harmed or traumatized by the event?
- Is this child able to participate in court proceedings?
- Other questions relevant to the specific case.

Evaluators should use multiple methods of data gathering.

The evaluator should be provided with certain background information, which includes:

- CP&P investigation report (or summary report) that is current/up to date;
- Existing prior psychological and psychiatric evaluations of the child and biological parent(s);
- Available law enforcement records including police reports; criminal charges and convictions; taped interviews, if available; and Promis/Gavel history of offenses;

- Prior CP&P history, including all prior referrals, with a finding for each allegation/investigation; investigative summaries;
- Complaint filed in court; and,
- Known mental health, substance abuse, or domestic violence history.

If a child is to be evaluated, the CP&P caseworker assigned to the case should accompany the child to the evaluation to support the child, to be available to provide any additional information and to hear directly from the evaluator any initial findings or recommendations. Whenever possible so as to best inform the evaluation, the investigative worker should accompany the child. Whenever possible a trusted adult should also accompany the child.

During the clinical interview, an evaluator:

- Establishes “Ground Rules” between the evaluator and the child.
- Explains to the child, in age appropriate and developmentally appropriate terms, the nature and the scope of the evaluation.
- Establishes the child’s developmental and cognitive ability to participate in the evaluation.
- Establishes the child’s competency. Does the child know the difference between the truth and a lie, real or pretend?
- Obtains the child’s version of the incident.
- Notes the child’s affect upon describing the incident.
- Asks questions to gather past history.
- Determines family relationships.
- Determines peer relationships.

Once the evaluation has been completed, the summary and report should include:

- Reason for the report – summary background;
- Nature of the allegation;
- Prior history;
- Documentation including a summary of the interview and direct quotes by the person being interviewed;
- Clinical finding and explanation;
- Any formal diagnosis;
- Clinical determination – indicate whether supported/not supported; and,
- Recommendations.

2. During Permanency Planning/Hearings

The Role and Function of Forensic Evaluations during Permanency Planning/Hearings

At the time of referral and over the course of a child or adolescent’s time under CP&P custody, mental health evaluations may be required to contribute to the decisions made about placement,

permanency, and parental rights. During permanency planning and hearings, evaluations are often used for:

- Interim Evaluation: The interim evaluation is meant to guide reasonable efforts for reunification.
- Ten Month Conference: The ten month conference is used to prepare a permanency plan for the child or youth in out-of-home placement. Before moving forward, any previous reports should be reviewed. It would be useful if the evaluator from the interim evaluation was also utilized at this point.
- Periodic Evaluation – Evaluation of Imminent Concerns Arising during Placement: An evaluation of imminent concerns is used to assess any risks or challenges that the child may incur during the course of the protective services or guardianship litigation. Examples include:
 - Disruption of the current placement;
 - Acute crisis (e.g., psychiatric hospitalization, severe medical illness, runaway, arrest, school disruption); and
 - Significant change in response to visitation.

Forensic Evaluation Process during Permanency Planning/Hearings

The purpose of the evaluation should be clear and should outline the specific questions to be addressed by the evaluation. The following referral questions should help to guide forensic evaluations at each of the stages identified for permanency planning/hearings:

Interim Evaluations Referral Questions:

- What services are needed for reunification?
- What impact has the abuse/neglect history had on the child?
- What are the risks that need to be addressed?
- Is the parent fit and able to parent the child?
- What actions are recommended to address the risks?
- What are the strengths that can be built upon?
- What visitation can be safely afforded between parents and their child(ren)?

Ten Month Conference Referral Questions:

- What progress has been made towards eliminating the harm?
- What still needs to be done?
- Are there any new areas of need?
- If a home other than the child(ren)'s current placement is being considered, is it in the best interest of the child(ren) to move to another placement if proposed by the parents, or to stay permanently where he or she is residing?

- Can this child transition back to the biological parents, without experiencing more harm than good?
 - If bonding and attachment are issues, an evaluation by a psychologist is necessary. A psychiatrist may contribute information within his or her area of expertise.
- Have the correct services been provided so far, and is there a need for a reduction, modification, or expansion of services?

It may be necessary to reevaluate the permanency plan. All of the questions above would apply to any such reevaluations.

Evaluators should use multiple methods of data gathering.

For these evaluations, the evaluator should be provided with certain background information, which includes:

- Existing prior psychological and psychiatric evaluations of the child and biological parent(s);
- Existing treatment reports for biological parents and child;
- Known mental health, substance abuse, or domestic violence history;
- Visitation reports;
- Complaint for guardianship, if filed;
- CP&P investigation report (or summary report) that is current/up to date;
- Prior CP&P history, including all prior referrals, with a finding for each allegation/investigation; investigative summaries;
- Most recent CP&P court report;
- Important selected contact sheets from the CP&P case record;
- Available law enforcement records including police reports; criminal charges and convictions; taped interviews, if available; and Promis/Gavel history of offenses;
- Additional information the parent wants to share with the evaluator; and,
- Any further available information requested by the evaluator.

All evaluations should include a review of comprehensive, accurate background information; a clinical interview; and the use of an appropriate assessment tool.

The evaluator should have access to all information he or she deems necessary in order to respond to the questions posed.

Periodic Evaluation – Evaluation of Imminent Concerns Arising during Placement Referral Questions:

- Identify impact of presenting problem.
- What are the recommended services or actions to address the problem?
- Should the permanency plan change?

For Periodic Evaluations of Imminent Concerns Arising during Placement, documented relevant information is needed as well as all available relevant reports, such as:

- Medical reports;
- Police reports;
- School reports;
- Psychiatric reports; and
- Relevant contact sheets.

3. During Litigation for Guardianship Complaints

The Role and Function of Forensic Evaluations during Litigation for Guardianship Complaints

Guardianship evaluations consist of fitness and bonding assessments during trial preparation after a guardianship complaint has been filed. Ideally, both the fitness and bonding assessments are completed by the same psychologist.

The presumption is that fitness and bonding assessments are required for guardianship litigation. It is recognized that in some circumstances, parent-child observations may not be necessary or advisable for purposes of the evaluation. For example, in cases where the safety or well-being of the child is clearly in jeopardy or parental contact with the child has been prohibited by a prior fitness and bonding assessment, parent-child observations may be bypassed. In such cases, the evaluator should note explicitly the reason(s) that a parent-child observation was not included.

A bonding evaluation assesses the relationship between the child(ren) and the proposed caregivers and other household members as appropriate.

Forensic Evaluation Process during Litigation of Guardianship Complaints

Guardianship Evaluation Referral Questions:

- What progress has been made towards eliminating the harm?
- What still needs to be done?
- Are there any new areas of need?
- If a home other than the child(ren)'s current placement is being considered, is it in the best interest of the child(ren) to move to another placement if proposed by the parents, or to stay permanently where he or she is residing?
- Can this child transition back to the biological parents, without experiencing more harm than good?
 - If bonding and attachment are issues, a psychological evaluation is necessary. A psychiatrist may contribute information within their area of expertise.
- Assess the child's bond and attachment to the biological parent(s).
- What harm, if any, will result if parental rights are terminated?

- Can the resource family parents mitigate the harm?
- Assess the child's bond and attachment to any proposed adoptive resource parent(s).
- Would severe and enduring harm occur if the child is removed from the proposed adoptive resource parents?
 - Can the biological parents mitigate the harm?

Guardianship evaluations call for specific competencies³ that are referred to in this section.

The evaluator at this stage in most circumstances will be a *licensed psychologist* or a psychiatrist. The licensed professional must be qualified to perform custody/parenting time evaluations and/or termination of parental rights evaluations through education, training, and/or supervision in all of the following categories:

1. Child growth and development;
2. Psychological testing;
3. Parent-child bonding;
4. Parenting skills;
5. Adult development and psychopathology;
6. Family functioning;
7. Child and family development;
8. Child and family psychopathology;
9. The impact of divorce or family dissolution on children; and,
10. The impact of age, gender, race, ethnicity, national origin, language, culture, religion, sexual orientation/identity, disability, and socioeconomic status on custody/parenting time evaluations.

When the following topics are involved, the licensed psychologist or psychiatrist shall have specialized education, training, and/or supervision in the specific topic, or the licensee shall refer to a licensed mental health care provider who has that education, experience, training, and/or supervision. The topic areas include:

1. Physical, sexual, or psychological abuse of spouse or children;
2. Physical and emotional neglect of children;
3. Alcohol or substance abuse that impairs the ability to parent;
4. Medical/physical/neurological impairment that affects the ability to parent; or
5. Other areas beyond the licensee's expertise that are relevant to the custody/parenting time evaluation.

Evaluators may identify relevant issues not anticipated in the referral questions that could enlarge the scope of the evaluation. At this stage, it is important to consider some relevant factors or issues in responding to the bonding and attachment referral questions.

These factors include:

³ N.J.A.C. 13:42-12.2. Available online at,
http://www.njconsumeraffairs.gov/adoption/psychado_090710.HTM

1. Age of the child;
2. The developmental stage of the child;
3. Child's history of abuse and/or neglect;
4. Child's resiliency;
5. Any special needs - medical or emotional - of the child or biological parents;
6. Parenting skills of both sets of parents;
7. Length of time in biological parents' care;
8. Number of placements;
9. Length of time in each placement;
10. Previous failed reunification attempts;
11. Child's wishes, weighted in accordance with developmental functioning;
12. Demonstrated willingness and ability of both biological parents and proposed adoptive resource parents to comply with services;
13. Demonstrated willingness and ability of both biological parents and proposed adoptive resource parents to recognize and meet the child's needs, including issues relating to reunification or adoption;
14. History of child's interaction with both biological parents and proposed adoptive resource parents;
15. Issues that may affect child's behavior during a bonding evaluation; and,
16. Sibling bonds/other attachments.

Evaluators should use multiple methods of data gathering.

Evaluators should be provided with the same background information listed under *Section 2: During Permanency Planning/Hearings*.

All evaluations should include a review of comprehensive, accurate background information; a clinical interview; and the use of an appropriate assessment tool.

The evaluator should have access to all information he or she deems necessary in order to respond to the questions posed.

Appendix B

The Quality Improvement Tool (QI Tool)

Evaluator/Evaluation Detail Sheet

Q1.1. Date of Review (MM/DD/YYYY)

Q1.2. Case Identification Number

Q1.3. Local Office Location

Q1.4. Evaluation conducted by a

Private Provider

☐

Regional Diagnostic and Treatment Center (RDTC)

☐

Q1.5. Reviewer

Q1.6. Is the DCP&P Case Goal stated in the Evaluation or Referral Page?

Yes

☐

No

☐

Q1.7. Reviewer Section Comments

Referral Information**Q2.1. Referral Information Obtained from SAR****Q2.2. Subject of Assessment**

Child

☐

Adult

☐**Q2.3. Subject of Assessment's Gender**☐ Male☐ Female☐ Transgender☐ Other☐ Not Specified**Q2.4. Date of Birth (MM/DD/YYYY)**☐ Specified☐ Not Specified**Q2.5. Indicate the Referral Question/Statement(s) by either selecting all that apply**☐

Assess current level(s) of cognitive functioning

☐

Service needs/ treatment

☐Assess current level(s) of psychological/ emotional/
behavioral functioning☐

recommendation(s) Other referral question/

☐Assess ability to provide adequate care and protection to
child/ parenting abilities.☐

statement

No Referral Question/Statement

Q2.6. If "Other" was selected for Q2 5 please write the referral question/ statement(s)

Q2.7. Referral questions/statements are clearly stated

Yes

☐

No

☐

Q2.8. If "No" was selected for question 2.6, "Referral questions/statements are clearly stated" briefly explain your selection.

Q2.9. What evaluation questions are addressed in the words of the Evaluator?

Case Detail Sheet

Q3.1. Purpose of Evaluation

- ☐ Investigation
- ☐ Permanency Planning / Hearing
- ☐ Litigation of Guardianship
- ☐ Other
- ☐ Not Specified / Unclear

Q3.2. Type of Evaluation

- ☐ Interim
- ☐ Ten Month
- ☐ Periodic
- ☐ Not Specified

Q3.3. Where is the child(ren) currently placed at the time of this evaluation?

- ☐ Biological Parent(s)

☐ Adoptive / Resource Parent(s)

☐ Other - please specify

Q3.4. Types of Assessment for Child (as indicated by the Evaluator)

☐ Abuse or Neglect

☐ Fire Setting

☐ Psychiatric

☐ Bonding

☐ Psychological

☐ Sexual Abuse

☐ Substance Abuse

☐ Other - Specify

Q3.5. Types of Assessment for Adult (as indicated by the Evaluator)

☐ Psychological

☐ Psychiatric

☐ Parenting Evaluation/Parental Capacity

☐ Bonding

☐ Substance Abuse

☐ Domestic Violence

☐ Sexual Abuse

☐ Other - Specify

Q3.6. Qualifications of Evaluator(s) (check all that apply)

☐ Licensed Psychologist - PhD / PsyD / EdD

☐ Psychiatrist

☐ MD / DO

☐ Licensed LPC

☐ LCSW

☐ Graduate Level Intern

☐ Other - Specify

Q3.7. Reviewer Section Comments

Background Information

Q4.1.

Interview Observations:

Instructions: Next to each item select: Yes - if the practice is described as part of the interview or No - if the practice is not described as part of the interview.

Q4.2. Subject of Assessment's Race

☐ Asian or Asian Indian

☐ Two or More Races

☐ Black or African American

☐ Other Race

☐ Pacific Islander

☐ Not Specified

☐ White

Q4.3. Is the Subject of Hispanic or Latino origin

Yes

☐

No

☐

N/A - Not Specified

☐

Q4.4. Does the Evaluator specify the ethnicity of the subject?

Yes

☐

No

☐

Q4.5. Does the Evaluator specify the primary language of the subject?

Yes

☐

No

☐

Q4.6. What language was specified?

☐ English

☐ French

☐ Spanish

☐ Arabic

☐ Chinese

☐ Russian

☐ Polish

☐ Hindi

☐ Korean

☐ Other

Q4.7. Does the Evaluator indicate how the evaluation was conducted?

☐ Through an Interpreter

☐ By the Evaluator - conducted using interviewee's primary language

☐

By the Evaluator - conducted using English

☐ N/A - Not specified

Q4.8. Does the Evaluator avoid using biased language?

Yes

☐

No

☐

Q4.9. Does the Evaluator avoid multiple relationships / conflicts of interests (Evaluator NOT providing therapy to interviewee or family member)?

Yes

☐

No

☐

Q4.10. Does the Evaluation contain direct quotes that describes the perceptions of the person being evaluated?

Yes

☐

No

☐

Q4.11. Reviewer Section Comments

Interview with Child

Q5.1. Was an interview with a child conducted?

Yes

☐

No

☐

Q5.2.

Interviews with Child

Instructions: Next to each item select: Yes - if the practice is described as part of the interview or No - if the practice is not described as part of the interview.

Q5.3. Does the Evaluator describe establishing any of the following with the child:

- ☐ Ground Rules of the Evaluation
- ☐ Nature and Scope of the Evaluation
- ☐ Competency

Q5.4. Does the Evaluator include the child's version of the situation?

Yes

☐

No

☐

Q5.5. Does the Evaluator describe the child's affect during the interview?

Yes

☐

No

☐

Q5.6. Does the Evaluator gather a history?

Yes

☐

No

☐

Q5.7. Does the Evaluator note and/or assess for any disabilities of the Child?

Yes - Disabilities Noted

☐

Yes - No Disabilities Noted

☐

No

☐

Q5.8. Does the Evaluator accommodate for any noted disabilities of the Child?

Yes

☐

Partial Accommodation

☐

No

☐

Q5.9. Does the Evaluator use age appropriate terms?

Yes

☐

No

☐

Q5.10. Does the Evaluator have the child describe any of the following relationships (check all that apply)?

Family Relationships

School Relationships

☐☐ Peer Relationships☐☐ Other Relationships - specify

Q5.11. Does the evaluator describe any other domains of the child?

Yes

☐

No

☐

Q5.12. If yes - please specify which other domains.

Q5.13. Reviewer Section Comments

Interview with Adult

Q6.1. Was an interview with an Adult conducted?

Yes

☐

No

☐

Q6.2. Is the Adult a(n):

☐ Biological Parent(s)☐ Adoptive / Resource Parent(s)☐ Other - please specify

Q6.3. Is the Adult the subject of the interview or is the interview a collateral interview?

Subject of the Interview

Collateral Interview



Q6.4.

Interviews with an Adult

Instructions: Next to each item select: Yes - if the practice is described as part of the interview or No - if the practice is not described as part of the interview.

Q6.5. Does the Evaluator describe establishing any of the following with the Adult?

- ☐ Ground rules of the evaluation
- ☐ Nature and scope of the evaluation
- ☐ Informed consent

Q6.6. Does the Evaluator describe the Adult's affect during the interview?

Yes



No



Q6.7. Does the Evaluator note and/or assess for any disabilities of the Adult?

Yes - Disabilities Noted



Yes - No Disabilities Noted



No



Q6.8. Does the Evaluator accommodate for any noted disabilities of the Adult?

Yes



Partial Accomodation



No



Q6.9. Does the Evaluator describe any of the following histories of the Adult (check all that apply)?

☐ Education History☐ Legal History☐ Work History☐ Medical History☐ Relationship History☐ Substance Abuse / Treatment History☐ Family History☐ Other History

Q6.10. Does the Evaluator describe the family relationships of the Adult?

Yes

☐

No

☐

Q6.11. Does the Evaluator describe the parenting style of the Adult?

Yes

☐

No

☐

Q6.12. Does the Evaluator describe the parenting capacity of the Adult?

Yes

☐

No

☐

Q6.13.

Instructions: The following questions are related to the current allegations.

Q6.14.

Does the Evaluator describe personal skills of the Adult?

Yes

☐

No

☐

Q6.15. Does the Evaluator describe risk factors of the Adult?

Yes

☐

No

☐

Q6.16. Does the Evaluator describe functional abilities of the Adult?

Yes

☐

No

☐

Q6.17.

Does the Evaluator take into consideration other relatives if or when a parent is unable to meet the needs of the child?

Yes

☐

No

☐

Q6.18. Reviewer Section Comments

Psychological Inventories and Interpretation

Q7.1. Were tools (psychological measures / scales / inventories) used to complete the evaluation?

- ☐ No tools were used; interview only.
- ☐ Tools were used.

Q7.2. Indicate which (if any) Cognitive and Achievement Test(s) were used.

- | | | |
|---|--|---|
| <input type="checkbox"/> Adaptive Behavior Assessment System (ABAS) | <input type="checkbox"/> Vineland Adaptive Behavior Scales | <input type="checkbox"/> Wechsler Preschool and Primary Scale of Intelligence (WPPSI) |
| <input type="checkbox"/> General Ability Measure for Adults (GAMA) | <input type="checkbox"/> Wechsler Abbreviated Scale of Intelligence (WASI) | <input type="checkbox"/> Wide Range Achievement Test (WRAT) |
| <input type="checkbox"/> Kaufman Assessment Battery for Children (KABC) | <input type="checkbox"/> Wechsler Adult Intelligence Scale (WAIS) | <input type="checkbox"/> Woodcock Johnson |
| <input type="checkbox"/> Kaufman Brief Intelligence Test (KBIT) | <input type="checkbox"/> Wechsler Individual Achievement Test (WIAT) | <input type="checkbox"/> Universal Nonverbal Intelligence Test |
| <input type="checkbox"/> Shipley Institute of Living Scale | <input type="checkbox"/> Wechsler Intelligence Scale for Children (WISC) | <input type="checkbox"/> Other (Specify) <input type="text"/> |
| <input type="checkbox"/> Stanford-Binet | <input type="checkbox"/> Wechsler Memory Scale (WMS) | <input type="checkbox"/> None |
| <input type="checkbox"/> Test of Nonverbal Intelligence (Toni) | | |

Q7.3. Indicate which (if any) Multiscale Inventories were used.

- | | |
|---|--|
| <input type="checkbox"/> Behavior Assessment System for Children (BASC) | <input type="checkbox"/> Personality Assessment Inventory (PAI) |
| <input type="checkbox"/> Child Behavior Checklist | <input type="checkbox"/> Personality Assessment Screener (PAS) |
| <input type="checkbox"/> Conners Behavior Rating | <input type="checkbox"/> Sixteen Personality Factors Questionnaire |
| <input type="checkbox"/> Millon Adolescent Personality Inventory (MAPI) | <input type="checkbox"/> Youth Self Report |
| <input type="checkbox"/> Millon Clinical Multiaxial Inventory (MCMI) | <input type="checkbox"/> Other (Specify) <input type="text"/> |
| <input type="checkbox"/> Minnesota Multiphasic Personality Inventory (MMPI) | <input type="checkbox"/> None |

Q7.4. Indicate which (if any) Clinical and/or Personality Scale(s) were used.

☐ Beck Depression Inventory

☐ Projective Drawings

☐ Beck Anxiety Inventory

☐ Psychiatric Diagnostic Screening Questionnaire

☐ Beck Hopelessness Scale

☐ Rorschach

☐ Beck Youth Inventory

☐ Sentence Completion

☐ Children's Depression Inventory

☐ Other (Specify)

☐ Conners ADHD

☐ None

Q7.5. Indicate which (if any) Psychological Tests in Child-Related Forensic Issues were used.

☐ Adult Adolescent Parenting Inventory

☐ Parent-Child Relationship Inventory

☐ Child Abuse Potential Inventory

☐ Parenting Stress Index

☐ Child Behavior Checklist

☐ Trauma Systems Checklist

☐ Child Sexual Behavior Inventory

☐ Other (Specify)

☐ Millon Adolescent Clinical Inventory

☐ None

Q7.6. Indicate if any Neuropsychological Tools were used.

Neuropsychology focuses on the relationship between brain functioning and behavior. Neuropsychological test batteries are generally broader in scope and more in-depth than traditional batteries, and hence provide a more thorough and detailed description of cognitive strengths and weaknesses.

Examples of neuropsychological tests:

Boston Naming Test

Delis-Kaplan Executive Function SYstem

Luria-Nebraska Neuropsychological Battery

Yes

☐

No

☐

Q7.7. Indicate and describe other tools that were used to complete the evaluation (excluding Neuropsychological Tools).

Q7.8. Does the Evaluator adequately describe the purpose of the psychological tools used in the Evaluation?

Yes

☐

No

☐

Q7.9.

Does the Evaluator describe how to interpret the results of the psychological tools used in the Evaluation?

Yes

☐

No

☐

Q7.10. Does the Evaluator use the current version(s) of the psychological tools used in the Evaluation?

Yes

☐

No

☐

Q7.11. Does the Evaluator describe any threats to the validity of the results (May select all that apply)?

☐ Yes☐ No☐ No, but threat(s) to validity are present

Q7.12. Indicate the threat(s) to the validity of the results (whether they were reported by the Evaluator or not).

☐ Issues pertaining to cultural factors☐ Issues pertaining to participant characteristics☐ Issues pertaining to test administration☐ Issues pertaining to validity indexes☐ Other threats to validity

Q7.13. Reviewer Section Comments

Evaluation Questions**Q8.1.**

Instructions: Next to each item select: Yes - if the practice is described as part of the interview or No - if the practice is not described as part of the interview.

Q8.2.

Does the Evaluator assess the risks that need to be addressed, as indicated by the referral?

Yes

☐

No

☐

N/A - Outside scope of Evaluation

☐

Q8.3. Does the Evaluator describe what progress has been made towards eliminating the risks / harm, as indicated by the referral?

Yes

☐

No

☐

Q8.4. Does the Evaluator identify the impact of the presenting problem, as indicated by the referral question (s), as indicated by the referral?

Yes

☐

No

☐

Q8.5. Does the Evaluator, within the scope of their professional judgment, identify the impact of the child's history of abuse / neglect?

Yes

☐

No

☐**Q8.6.**

Does the Evaluator identify additional steps needed to address the risks / harm?

Yes

☐

No

☐

Q8.7. Does the Evaluator identify areas of strength?

Yes

☐

No

☐

Q8.8. Does the Evaluator identify new areas of need, not previously identified?

Yes

☐

No

☐

Q8.9. If yes - please briefly list the new areas of need.

Q8.10. Does the Evaluator list and/or describe services that are currently provided?

Yes

☐

No

☐

N/A - Services are not currently provided

☐

Q8.11. Reviewer Section Comments

Collateral Information

Q9.1.

Collateral Information - Was the following information accessed by the Evaluator or were attempts made to access?

Instructions: Please check off next to each item if the Interview, Record, or History was accessed by the Evaluator.

Q9.2. Which of the following Collateral Interviews were accessed by the Evaluator?

☐ Interview with Child

☐ Interview with Parent(s)

☐ Interview with Spouse / Partner

☐ Interview with other Relative(s)

☐ Interview with Resource Family

☐ Interview with DCP&P Worker(s)

☐ Interview with Teacher(s)

☐ Taped Interview(s)

☐ Interview with other collateral - please specify

Q9.3. Which of the following Collateral Records and/or Histories were accessed by the Evaluator?

☐ Medical Records

☐ Court Complaint

☐ CP&P Records

☐ Mental Health / Psychiatric Records

☐ School Records

☐ Substance Abuse Records

☐ Prior Evaluations and/or Observations

☐ Domestic Violence Records

☐ Law Enforcement Records

☐ Prior Complaints Filed

☐ History of Offenses

☐ Other Records - please specify

Q9.4. To what extent were appropriate background materials provided by CP&P to address the purpose of the referral

No background material(s) listed

☐

Insufficient

☐

Sufficient

☐

Q9.5. Please explain your previous selection

**Q9.6.
Reviewer SectionComments**

Diagnosis

Q10.1. Does the Evaluator conduct a diagnostic assessment on the Subject of the Evaluation (including exploration or discussion of possible diagnoses)?

Yes

☐

No

☐

Q10.2. Does the Evaluator provide a DSM diagnosis (or ICD 10) for the Subject of the Evaluation?

☐ Yes, by the Evaluator

☐ Yes, by History

☐ No

Q10.3. Was there reasonably sufficient evidence or basis for the diagnosis made?

Yes

☐

No

☐

Q10.4. Should the Evaluator have conducted a diagnostic assessment?

☐ Yes

☐ No

☐ N/A - Termination of Parental Rights Case

Q10.5. If yes, please explain why.

Q10.6. Reviewer Section Comments

Summary

Q11.1. Was a summary included?

Yes

☐

No

☐

Q11.2.

Summary of Findings

Instructions: Next to each item select: Yes - if the practice is described as part of the interview or No - if the practice is not described as part of the interview.

Q11.3. Does the Evaluator provide a background summary?

Yes
☐

No
☐

Q11.4. Does the summary address referral questions?

Yes
☐

No
☐

Q11.5. Does the Evaluator describe the nature of the allegation(s)?

Yes
☐

No
☐

Q11.6. If applicable, does the Evaluator indicate clinical findings in the summary?

Yes
☐

No
☐

N/A
☐

Q11.7. Does the Evaluator identify clinical interventions?

Yes
☐

No
☐

Q11.8. Reviewer Section Comments

Recommendations

Q12.1. Were recommendations made?

Yes

☐

No

☐

Q12.2.

Recommendations

Instructions: Next to each item select: Yes - if the practice is described as part of the interview or No - if the practice is not described as part of the interview.

Q12.3. Does the Evaluator make recommendations to address the needs of the child described in the referral, regardless of the subject of the evaluation?

Yes

☐

No

☐

Q12.4. Does the Evaluator make recommendations that were relevant to the purpose of DCP&P?

Yes

☐

No

☐

Q12.5. Does the Evaluator make recommendations to address the risk and/or harm that include:

- ☐ Services that are evidence based
- ☐ Services that are generally accepted in clinical practice as appropriate for use
- ☐ Services that are not evidence based, but are evidence informed or promising programs
- ☐ Services for which there is either weak or no evidence
- ☐ It is uncertain whether some of the recommendations are supported by current scientific evidence N/A
- ☐ - Noclinical or program based service recommendations

Q12.6. Does the Evaluator make recommendations that include a description of expected outcomes from interventions and/or treatments?

Yes

☐

No

☐

Q12.7. Does the Evaluator make recommendations that describe conditions most likely to yield successful outcomes?

Yes

☐

No

☐

Q12.8. Does the Evaluator include a hierarchy timeline for service based recommendations to be completed?

Yes

☐

No

☐

Q12.9. If the Evaluator included a hierarchy timeline for service based recommendations do they indicate which services are to occur concurrently?

Yes

☐

No

☐

Q12.10. Does the Evaluator make recommendations that describe the degree to which specific interventions/treatment are likely to be successful?

Yes

☐

No

☐

Q12.11. Does the Evaluator make recommendations that take into account the individual's race, ethnicity, gender or other self-defining characteristics?

Yes

☐

No

☐

Q12.12. How many clinical or program based service recommendations were specified by the Evaluator?

0

☐

1-2

☐

3-5

☐

> 5

☐

Q12.13. How many other service recommendations were specified by the Evaluator?

0

☐

1-2

☐

3-5

☐

> 5

☐

Q12.14. Reviewer Section Comments

Evaluation Rubric

Q13.1.

Referral Statements:

#{q://QID12/ChoiceTextEntryValue/1}

#{q://QID12/ChoiceTextEntryValue/2}

#{q://QID12/ChoiceTextEntryValue/3}

#{q://QID12/ChoiceTextEntryValue/4}

#{q://QID12/ChoiceTextEntryValue/5}

Q13.2.

Addressed Evaluation Questions:

#{q://QID20/ChoiceTextEntryValue}

Q13.3.

Interview with a child:

#{q://QID49/ChoiceTextEntryValue}

Q13.4. Interview with an adult:

#{q://QID66/ChoiceTextEntryValue}

Q13.5. Tests and measurements:

#{q://QID82/ChoiceTextEntryValue}

Q13.6. Evaluation Questions:

#{q://QID103/ChoiceTextEntryValue}

Q13.7.

Bonding and Attachment Section Reviewer Comments:

#{q://QID146/ChoiceTextEntryValue}

Q13.8. Summary Section Reviewer Comments:

`{q://QID156/ChoiceTextEntryValue}`

Q13.9.

Recommendation Section Reviewer Comments:

`{q://QID168/ChoiceTextEntryValue}`

Q13.10. Instructions:

For the following questions please indicate the degree to which the evaluation/evaluator meets the following criteria, using the scale: 1 being Strongly Disagree and 6 being Strongly Agree.

Q13.11. To what degree does the evaluation contain relevant, professionally sounds observations, results, and opinions.

1 2 3 4 5 6

○ ○ ○ ○ ○ ○

Q13.12. To what degree does the evaluation address the particular psychological, behavioral, and developmental needs of the child and/or parent(s).

Q13.13. Does the Evaluator describe any of the following:

- ☐ Degree to which severe and enduring harm would occur if the child is removed from their current placement
- ☐ Degree to which the parent(s) are fit and able to parent the child
- ☐ Degree to which the parent(s) can mitigate harm
- ☐ Degree of the relationship between the child and the parent(s)
- ☐ Making recommendations to change the permanency plan
- ☐ Appropriate visitation between the parent(s) and the child
- ☐ Services needed for reunification

Q13.14. To what degree does the Evaluator use a multimodal approach to draw conclusions about the current case.

1 2 3 4 5 6



- ☐ Cultural explanations of the allegations
- ☐ Client's cultural background and ways it might influence their behavior
- ☐ Client's cultural background and ways it might influence their parenting capacity

1 2 3 4 5 6

○ ○ ○ ○ ○ ○

1 2 3 4 5 6

○ ○ ○ ○ ○ ○

1 2 3 4 5 6



Q13.22. Please indicate the overall quality of the forensic evaluation written by the Evaluator.

1
☐

2
☐

3
☐

4
☐

5
☐

6
☐

Q13.23. Is a secondary review of this case needed?

Yes
☐

No
☐

Q13.24. Reviewer Section Comments

Assessments Pilot Study

Q14.1. The following section is for for exploratory purposes.

Q14.2. Given the referral question(s) and the background information for this case, would you make any changes to the battery of psychological tests administered?

Yes
☐

No
☐

Q14.3. Should the Evaluator have (May select all that apply):

Added Tests
☐

Omitted Tests
☐

Replaced Tests
☐

Q14.4. What test(s) should have been added?

Q14.5. What test(s) should have been omitted?

Q14.6. What test(s) should have been replaced, and with what?

Q14.7. If these changes were made, what impact do you think this would have on the overall quality of the evaluation?

- ☐ A great deal
- ☐ A lot
- ☐ A moderate amount
- ☐ A little
- ☐ None at all

Q14.8. Please explain your response to the previous question