6-2019

Defensiveness in Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF) Scores of Physically Abusive Parents

Julia Meehan Moroney

Defensiveness in Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF) Scores of Physically Abusive Parents

By

Julia Meehan Moroney, M.S.

Clinical Psychology

Florida Institute of Technology

A doctoral research project submitted to Florida Institute of Technology in partial fulfillment of the requirements for the degree of Doctor of Psychology in Clinical Psychology

Melbourne Florida

June 2019

We, the undersigned committee, hereby approve the attached Doctoral Research Project

Defensiveness in Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF) Scores of Physically Abusive Parents

By

Julia Meehan Moroney, M.S.

_____
Radhika Krishnamurthy, Psy.D., ABAP
Professor, Clinical Psychology
Committee Chair

_____
Travis Conradt, Ph.D.
Assistant Professor of Psychology

_____
John Lavelle, Ph.D
Associate Professor of Arts and Communication

_____
Lisa Steelman, Ph.D.
Dean
College of Psychology and Liberal Arts

Defensiveness in Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF) Scores of Physically Abusive Parents

Julia Meehan Moroney, M.S.

Major Advisor: Radhika Krishnamurthy, Psy.D., ABAP

Abstract

There is a paucity of research on the personality test profiles of physically abusive parents. Given that personality assessment is typically a major component of dependency evaluations, the lack of studies in this area represents a significant gap in understanding personality-based factors contributing to physical abuse perpetration. Considering research findings of high levels of defensiveness found in profiles of parents undergoing child custody evaluations, it is reasonable to expect the same for parents involved in substantiated cases of child physical abuse. The current research was designed to examine the level of defensiveness in Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) profiles of physically abusive parents. Subsequently, the current research aimed to establish optimal cutting scores for the MMPI-2-RF Restructured Clinical (RC) scales for physically abusive parents to adjust for defensiveness and denial of personal problems. The central sample for the study consisted of 62 parents who had a substantiated allegation of child physical abuse. This sample was expected to respond defensively via elevated scores on scales Lie (L-r) and Correction (K-r). This hypothesis was largely confirmed, particularly in terms of high L-r scores. It was also hypothesized that RC

scale scores would be suppressed for the Physical Abuse sample, which was found for the majority of RC scales, with some exceptions. The study also employed two comparison samples: a nonabusive child custody sample (*N*=64) and a nonabusive community sample (*N*=61). Comparison between the three groups indicated the Physical Abuse group was the most defensive, most significantly differentiated by L-r. In keeping with the primary goal of the study, Receiver Operating Characteristics (ROC) analyses were conducted to establish optimal cutting scores for the RC scales for physically abusive parents. These were derived from score comparisons of the Physical Abuse and Community samples. Optimal cutting scores ranged from T=47 to T=59, and were predominantly at or below the MMPI-2-RF normative mean. These optimal cutting scores are intended for use in evaluations to adjust for defensive responding among physically abusive parents. Implications of these findings are discussed.

Keywords: MMPI-2-RF, physically abusive parents, child custody

Table of Contents

List of Tables

Introduction

Five-year-old Jessica was in her kindergarten class when her teacher noticed she was moving slowly and not playing with the other children as she usually did. Complaining of stomach pain, Jessica was sent to the nurse's office. Upon inspection, the nurse saw deep bruising around her abdomen, and further examination revealed bruises in varying states of healing covering her arms and legs. Jessica's mother and stepfather had been hitting her with objects and occasionally kicking her as punishment for behaviors such as leaving toys out or bedwetting. Sadly, cases like Jessica's are not uncommon. In 2005, physical abuse was the second most frequently substantiated form of maltreatment suffered by children in the United States, and 16.6% of United States children have been confirmed victims of such abuse (USDHHS, 2007). This number does not include those children whose abuse has not been reported, or whose cases were unable to be substantiated (USDHHS, 2007). Additionally, data from the Child File, a national database that records information about child abuse and neglect, has revealed that the younger a child is the more at risk he or she is for maltreatment (American SPCC, 2016).

The welfare of children is of great concern to society. Since the establishment of the Society for the Protection and Care of Children (SPCC) in 1875, concern for child wellbeing has been demonstrated through the presence of government-based child protection services in every state, such as Florida's Department of Children and Families (DCF) created in 1996. Additionally, this concern can be seen through laws created specifically with the intention of protecting children from harm both in the

immediate future, such as laws regarding child abuse and neglect, and the distant future, such as compulsory student education laws requiring children to receive adequate schooling. Florida Statute 827.03 regarding child abuse and child neglect describes physical abuse as a willful act that inflicts physical harm upon a child. It also includes the term *aggravated child abuse*, which describes a higher level of severity of such an act. The definitions of child abuse as outlined in the statutes of other states differ slightly, although they tend to include a willful act that results in physical harm to a child (Texas Family Code, 2005; New York Penal Law, 2015; Child Abuse and Neglect Reporting Act, 1987).

On such occasions where it is determined the child is at substantial risk for future harm and needs to be removed from his or her home, the child's status is designated as dependent (Proceedings Related to Children, 2005), that is, under the jurisdiction of the state for appropriate placement and care. The child may be placed with another relative or sheltered in a foster home until goals for a permanency plan are met or parental rights are terminated (Department of Children and Families, 2012). In instances such as these, one or both parents of the child may be required to submit to a psychological evaluation to help determine parental fitness. These evaluations typically include an interview with each individual, personality assessment, and evaluation of parenting ability. Parents undergoing this process have a great deal at risk during this time, such as losing their parental rights. Because of this, they often try to minimize their difficulties and shortcomings in order to appear psychologically healthy and well-adjusted. Impression management such as this is often evident in the

results of their testing, and is referred to as defensive responding (Bagby, Nicholson, Buis, & Radovanovic, 1999).

Defensive responding creates problems in personality assessment as it can cause the results to be invalid or yield less accurate results that limit interpretive ability. For example, defensive responding in personality tests such as the Minnesota Multiphasic Personality Inventory-2 Restructured Form (MMPI-2-RF) that measure relatively stable characteristics of a person may yield a profile that is not only unrepresentative of the examinee's way of thinking and behaving, but is also an inaccurate portrayal of his or her psychological adjustment. Profiles that reflect a respondent's actual psychological health and adjustment are especially important to obtain in evaluations conducted in a legal context due to their contribution to high-stakes decisions. Attempts have been made to gauge the level of defensiveness in respondents' profiles through the use of validity scales, and some studies have found that the use of cutting scores and statistical adjustments can be helpful in adjusting for defensive responding while simultaneously preserving the interpretive integrity of the measure. Problematically, however, there is a notable imbalance in the content of the maltreatment-related research, namely that much of the research is centered on child sexual abuse or all types of child maltreatment collectively as a single variable, and physical abuse-specific literature is markedly absent. The study of physical abuse is highly important as it is a very consequential issue for children. Children have the potential to be adversely impacted in many areas of their lives when they suffer physical abuse, including their physical and mental health, the way they raise their

own children, their choice of romantic partners, and their likelihood of substance use (Springer, Sheridan, Kuo, & Carnes, 2007). Moreover, basic physical safety is considered a human right in the United States, and all vulnerable populations, including children, are entitled to it. Due to the lack of research on such a significant issue, the current study pursued the identification of optimal cutting scores for MMPI-2-RF profiles of physically abusive parents in order to adjust for defensive response patterns.

Review of Literature

Characteristics of Physically Abusive Parents

Researchers examining why child physical abuse occurs consistently point to the combination of multiple contributing factors leading to abuse perpetration. These factors often include, but are not limited to, characteristics of the perpetrator, child characteristics, perpetrator-victim relationship, cultural practices, and economic stress (Berkout & Kolko, 2016; Herrmann & Martin, 1988; Schnitzer & Weigman, 2005). Attempts have been made to develop abuser typologies based on their shared characteristics, which can be used to more easily assess the interactions between abuser characteristics and the aforementioned factors (Francis, Hughes, & Hitz, 1992). The examination of characteristics of physically abusive parents can also aid in early identification, improvement of prevention programs, and improvement of treatment programs (Perry, Wells, & Doran, 1983), particularly when considering that abusive parents may have higher-than-average level of family conflict and low level of family cohesion occurring within their home (Perry et al., 1983; Stith et al., 2009).

Studies have shown that physically abusive parents are a heterogeneous group and vary in personality type and features (Francis et al., 1992), thus necessitating an alternative approach, such as identifying subtypes, to understand personality factors contributing to abusive actions. This was done in a cluster-analytic study by Francis et al. (1992) utilizing the Sixteen Personality Factor Questionnaire (16-PF). These researchers subdivided their sample of physically abusive parents into five groups based on 16-PF personality characteristics. The first group consisted of parents who

were shy, felt guilty and insecure, and tended to isolate themselves from others. The authors noted that this cluster reflects a combination of two typologies often found in the literature with regard to demographic and personality characteristics. This cluster pattern was similar to the findings of Stith and colleagues (2009), in which low self-esteem, depression, psychopathology, childhood abuse, and social isolation were common characteristics among physically abusive parents. Additionally, Milner and Chilamkurti (1991) had noted from the previous literature that perpetrators have low self-esteem, which appears to influence parent perceptions of child behavior as well as the parent's ability to manage stress, and depression was linked to parent-perpetrated physical abuse of a child although the nature of that link has yet to be clarified by research (Milner & Chilamkurti, 1991). Parents in Francis et al.'s (1992) second cluster produced personality profiles similar to what is expected for people in the general population. Parents in this group also tended to have higher levels of education and fewer children. Parents in the third cluster were similar to typologies found in the previous literature in terms of being socially skilled but compulsive. They tended to deny their pathology and made substantial efforts to present themselves in a favorable light (Francis et al., 1992). This is concordant with literature that has suggested perpetrators have an external locus of control and tend to blame their problems on outside forces such as other people and chance (Milner & Chilamkurti, 1991). Parents who were classified into the fourth cluster were noted to be compulsive in their behavior, and passive and submissive in their relationships with others. The partners of the parents in this cluster were also often involved in the commission of the abuse.

Parents in the fifth cluster established by Francis and colleagues (1992) were also socially withdrawn and isolated and tended toward tension, suspicion of others, and feelings of apprehension and frustration. Unlike those of the first cluster, however, the profiles produced by these parents indicated emotional lability and more severe psychological disturbance than any other cluster in the study. Contrary to the shy, withdrawn type of abuser in this cluster, other studies have found an abuser type where perpetrators of physical abuse tend to be more physically and verbally aggressive in interactions with others. Perpetrators who fit into this type also tend to have more negative interactions with their children but tend to interact less with their children overall compared to nonabusive parents. They turn to physical and punitive means of controlling their children rather than reasoning, and tend to perceive those strategies as being more effective (Milner & Chilamkurti, 1991).

Perry et al. (1983) noted that one of the only gender differences in perpetrators' psychological characteristics was that the male perpetrators in their sample did not exhibit lower self-esteem or higher anxiety compared to their non-abusing counterparts in the control group. They hypothesized that their abusive behaviors may be influenced more by situational and family-related factors, while recognizing that personality factors contributing to physically abusive actions may not have been fully considered in their study. Personality features often seen across studies include rigidity, loneliness, difficulty forming attachment to others, less empathy for others, more anger and less assertiveness, and higher levels of anxiety (Milner & Chilamkurti, 1991).

Studies have also attempted to shed light on the role of parenting stress, parenting capacity, and related factors that affect parent-child interactions. For example, Berkout and Kolko (2016) conducted a study with a sample of maltreating parents that specifically examined the effects of parenting stress, negative affect, and positive parenting behavior using results from factor-specific measures. Their goal was to determine the effects of various combinations of these factors on the commission and severity of physical abuse. Using a structural equation model, the researchers found significant indirect effects of parenting stress on physical abuse. According to their model, parenting stress predicted exacerbation of negative affect, which in turn was related to child-directed aggression. Stith and colleagues' (2009) findings support this model. They noted abusive parents tended to react more intensely to child-related aversive stimuli. Whipple and Webster-Stratton (1991) found a similar relationship between parental stress, negative affect, and physical abuse. In their study of abusive mothers they noted that those who were experiencing stress and clinical depression displayed higher levels of irritability as well as a reduced ability to manage their negative emotions.

Some researchers postulate certain biological factors related to parents' reactions to stress could contribute to their proclivity to abuse their children. For example, studies have often found that perpetrators perceived a greater impact of stress than non-abusers (Perry et al., 1983). This is not to say that abusive parents necessarily experience a higher degree of stress compared to non-abusive parents. Rather, it has been suggested that they experience more intense physiological

reactions to stress compared to nonabusers and lack the psychological resources and resilience to manage their experienced stress (Milner & Chilamkurti, 1991). Specifically, many researchers investigating physiological reactivity posit that perpetrators of child physical abuse display more intense autonomic reactions to stress and remain autonomically aroused for longer periods of time compared to non-abusers (Milner & Chilamkurti, 1991).

A study by Herrmann and Martin (1988) found the age of the child to be a significant predictor of hospitalization resulting from having been physically abused; specifically, children under the age of 3 were more likely to be hospitalized. They noted that the age of the child was not directly responsible for this finding. Rather, they theorized that, in addition to the physical vulnerability of children in this age group, more parenting skills are required to manage children of this age group's egocentrism and lack of communicative abilities. Those who are physically abusive may lack the requisite parenting skills, thereby necessitating examination of parent characteristics in addition to other factors. This is consistent with Berkout and Kolko's (2016) study that found parents who reported physically abusing their children also tended to report poorer parenting practices overall, such as less involvement with their children, inconsistently following up with consequences, and poorer monitoring of child behavior. The authors suggested these parents felt they had fewer methods with which they could manage their children's behavior, and thus resorted to physically abusive tactics (Berkout & Kolko, 2016). Many researchers posit that some cognitive deficits can contribute to child physical abuse by limiting the perpetrator's ability to

think flexibly and understand his or her child's behavior or come up with appropriate behavior management ideas. Additionally, researchers have hypothesized that cognitive deficits affecting communication may negatively impact an abuser's ability to effectively express their needs and cope with family difficulties (Milner & Chilamkurti, 1991). Studies have also found abusive parents to report experiencing a greater amount of physical maladies compared to non-abusers (as reviewed by Milner & Chilamkurti, 1991). Parents with more physical ailments may be less capable of effectively using consistent, nonabusive strategies due to fatigue, pain, and stress from their illnesses.

Parental perception of the child's behavior, rather than the child's actual misbehavior, has also been shown to be closely related to perpetration of physical abuse (Stith et al., 2009). Studies have shown that abusive parents, especially abusive mothers, tend to perceive their children as intentionally disobedient and often view their children more negatively compared to nonabusive parents. Additionally, they tend to ignore positive behaviors and often see negative behaviors as being more offensive compared to nonabusive parents (Milner & Chilamkurti, 1991; Stith et al., 2009). Combined with the strength of the relationship between parental anger and hyper-reactivity to stress that has previously been discussed, this finding suggests parental characteristics play a large part in the perpetration of child physical abuse (Stith et al., 2009). Parental expectations for their children have also been shown to be influential. Previous studies have found that abusive parents have unrealistically high expectations for their children, but more recent literature suggests abusive parents'

expectations are often inappropriately low regarding some areas and inappropriately high regarding others (Milner & Chilamkurti, 1991). A study by Perry et al. (1983) found that abusive parents expected their children to reach developmental milestones later than normal, contrasting with previous literature suggesting that parents expect children to develop more quickly than normal. The abused children in this study tended to display developmental delays, which the authors suggested may have impacted the parents' reported expectations for their children's development. However, the researchers also hypothesized that some of the abusive parents may have infantilized their children, leading to a lack of the requisite support for normal growth and development. Perry and colleagues (1883) posited that child expectations for independence clashing with parental expectations regarding child development may contribute to abuse (Perry et al., 1983). These finding emphasize the importance of continuously examining parental expectations and perceptions of children at different developmental levels.

Researchers acknowledge that a substantial amount of studies have focused on parental pathology when analyzing factors that contribute to perpetration of physical abuse. However, they also emphasize that because abuse occurs within the context of the family, family factors are also important to examine (Black, Heyman, & Slep, 2001; Stith et al. 2009). Some demographic factors have been linked to negative traits that contribute to later perpetration of physical abuse, although these factors are often debated amongst researchers (Milner & Chilamkurti, 1991). Research has suggested that indicators of lower socioeconomic status (SES) such as lower educational level,

lower income, and single parent status contribute to the cycle of abuse (Milner & Chilamkurti, 1991). In contrast, Schnitzer and Ewigman (2005) examining instances of fatal child maltreatment discovered that there was no increased risk to children in single-parent homes unless an unrelated adult (e.g., stepfather or boyfriend of the mother), also lived in the home. In these cases, the risk of death to children under the age of 5 who were dwelling in a home with adults unrelated to them was approximately 50 times higher than children who lived with two biological parents (Schnitzer & Ewigman, 2005).

Francis et al. (1992) also found variations in education level, number of children in the home, and parental age, suggesting none of these factors are solely influential on the parents' physically abusive acts (Francis et al., 1992). Further support for this comes from Stith and colleagues (2009) and their finding that the influence of parental age is often overshadowed by other mitigating and aggravating factors, as well as from Berkout and Kolko's (2016) finding of no relationship between race and perpetration of physical abuse. Berkout and Kolko (2016) posited that other variables may be at work, such as lack of resources or negative parental affect resulting from perceived racial discrimination. Perry and colleagues' (1983) study, mentioned earlier, further examined the influence of demographic factors found among physically abusing parents, using matched controls. In this study participants were matched based on social class, one or two parents in the home, age and sex of the child victimized, age of the parents, number of family members, and birth order of the victimized child. Perry et al. (1983) found evidence that abusive childhoods did not

determine abusive behaviors in adulthood, in contrast to the findings of Francis et al. (1992) that suggested an abusive childhood may impact personality factors seen in one of the abuser subtypes. In studies examining commonalities in cases of maltreatment-related fatalities, researchers have found that the biological father was most often the perpetrator, followed closely by unrelated father figures such as stepfathers or boyfriends of the biological mothers (Klevens & Leeb, 2010). Overall, men tended to perpetrate physical abuse significantly more often than women, a finding that is consistent with previous literature (Damashek, Nelson, & Bonner, 2013; Klevens & Leeb, 2010; Schnitzer & Ewigman, 2005).

Evaluation of Physically Abusive Parents

Much of the research on evaluations of parents who have physically abused their children is embedded within the literature on parental evaluations conducted within a range of contexts, including child custody evaluations, dependency cases, and parenting capacity evaluations, where physical abuse has not occurred. Such evaluations are utilized to aid in treatment planning, assess the efficacy of treatment, screen for signs of maltreatment or adjustment difficulties, confirm report contents, and predict recidivism in instances of maltreatment (Milner & Murphy, 1995). Many practitioners have found that psychological assessment is necessary in order to screen parents who are considered at risk of maltreating their children, identify those who have abused their children, and evaluate methods of abuse prevention and intervention (Milner, 1991). Following the confirmation that physical abuse has occurred, the

parent's level of risk for recidivism is assessed to determine whether the child is safe in his or her care (Milner, 1996).

One of many challenges faced by evaluators is in trying to predict future perpetration of physical abuse when deciding the best interest of the child in terms of placement (Milner, 1996). Models aimed at determining why child physical abuse occurs have been developed to help predict this risk (Milner, 1996). These models can serve to provide a framework within which evaluations are conducted and guide evaluators in their determinations of tools to use in the evaluation, factors present that contribute to abuse potential, and questions to answer about the family dynamics. However, researchers are often skeptical that these models possess adequate empirical support to allow them to be in accordance with evidentiary laws (Mart, 2003). Contributing to this is the fact that operationalization of physical abuse has historically been problematic and there is often overlap with other forms of maltreatment in research studies (Milner, 1996). This makes it harder to determine risk consistently and evaluate risk assessment methods across studies (Milner, 1996).

The guidelines set out by the American Psychological Association (2010) regarding child custody evaluations mandate the use of more than one source of information when evaluating parental risk for continued abuse. Such sources include psychological testing, clinical interviewing, behavioral observation, and a review of records when appropriate. Considering the outcomes of child custody evaluations have high stakes, it is critical that psychologists utilize methods that strengthen the reliability and validity of their conclusions (American Psychological Association,

2010). Additionally, some authors have suggested that practitioners conduct multiple evaluations on different dates to account for additional variables that may come into play after the first evaluation (Milner, 1996). Although the guidelines of the American Psychological Association are applicable across the country, the precise criteria considered in an evaluation may differ between states. Some examples include the ability of the involved parties to meet the basic needs of the child, moral fitness of the involved parties, the relationship between the child and involved parties and their ability to provide the child with love and affection, and the mental and physical health of the involved parties (Mart, 2003).

Various measures are employed to assess multiple areas of the parents' functioning, including measures that have been developed to assess specific risk factors associated with child physical abuse (Milner & Murphy, 1995). Specific parenting questionnaires are frequently used in child custody evaluations, such as the Bricklin Perceptual Scales (BPS), which is unique, albeit limited in scope in that it relies on a child's report of the parent's competence in various areas. The Ackerman-Shoendorf Scales for Parent Evaluation of Custody (ASPECT), like the BPS, was created specifically for use in child custody evaluations, and incorporates the results of other measures taken throughout the course of the evaluation (Jaffe & Mandeleew, 2011). One measure specifically developed to screen for child physical abuse is the Child Abuse Potential Inventory (CAPI; Milner & Murphy, 1995). This measure contains 160 items and is designed to obtain objective information on personality and other characteristics to help determine a parent's risk level for child physical abuse by

assessing for the presence of characteristics consistent with those of abusive parents (Milner, 1991; 1996). A high overall score on this measure indicates a greater risk of the examinee abusing his or her child. The characteristics are organized into six factor scales, including distress, rigidity, unhappiness, problems with the child and self, problems with the family, and problems with others. The CAPI also includes several validity scales, including Lie, Random Response, and Inconsistency, which are designed to detect overreporting, underreporting, and random response sets. Although the CAPI cannot be used by itself to identify abusive parents, studies assessing its use and psychometric qualities have noted it often revealed difficulties in parent-child interactions and was typically effective in discriminating between abusive and nonabusive matched comparison groups. Studies have also suggested that this test is best used with physically abusive parents due to its measured constructs; however, some studies have noted that perpetrators of sexual abuse and neglect often score higher than comparison groups (Heinze & Grisso, 1996). Despite its usefulness, it is limited because it may not apply to non-parent caretakers (Milner, 1996). Additionally, when demographics regarding type of injury and childhood illness come into play, the accuracy of results tends to become distorted (Milner & Murphy, 1995).

The Parenting Stress Inventory (PSI) is a screening tool designed to assess the level of stress related to childrearing experienced by a parent with a child under the age of 12. Like many similar measures, it has undergone several revisions since its development. The current fourth edition contains 120 items that load onto three domains of stressors, consisting of child characteristics, parent characteristics, and

situational/demographic life stress. This heavily-researched measure distinguishes stress factors related to parenting and child-rearing from general stress experienced by the respondent (Milner, 1991; 1996; Milner & Murphy, 1995). Although it does not distinguish abusive from nonabusive parents, it identifies problem areas that may be risk factors for abuse. This measure is typically used in the development of a treatment plan and later evaluation of the implemented plan (Heinze & Grisso, 1996; PAR Inc., 2012). The Michigan Screening Profile of Parenting (MSPP) is another measure that was developed as a screener for child physical abuse; however, investigations into its validity revealed it to be a more effective screener for difficulties in parent-child interactions (Milner, 1991; Milner, 1996). This self-report screener has shown distinctive differences between perpetrators and non-perpetrators, but it is not useful for definitively classifying parents into these groups (Milner & Murphy, 1995).

The Conflict Tactics Scale (CTS) is a measure that examines the types of reactions to conflict (Milner, 1991). Although the CTS contains a physical abuse scale, it has not been validated by research. Thus, its use remains limited to assessment of conflict reactions (Milner, 1996). Similarly, the Adult/Adolescent Parenting Inventory (AAPI) evaluates the expectations held by both parents and adolescents to gather information about their interactions (Milner, 1996).

In addition to parenting-related self-report measures, tests that examine respondents' emotional wellbeing are often included in parenting evaluation batteries. The State-Trait Anger Expression Inventory-2 (STAXI-2) measures three main aspects of the examinee's experience with anger, which are state anger, trait anger, and anger

expression. Originally, this measure was designed to be used to detect those who suppress their anger and identify ways of managing anger that could contribute to physical health problems. However, it has also been used to determine maladaptive ways of managing anger among abusive parents. The State Anger scale assesses the examinee's level of anger at a particular point in time, whereas the Trait Anger assesses the frequency with which feelings of anger occur over time. Anger Expression is separated into two main scales, Anger Expression and Anger Control. The Anger Expression scales measure the extent to which the examinee expresses feelings of anger to others, or contains feelings of anger in an unhealthy way. The Anger Control scales measure the examinee's control of feelings of anger through prevention of expressing such feelings toward people or objects, or controls feelings of anger through use of calming techniques. It also includes a total score, the Anger Expression Index, that provides an overall gauge of expression of anger (PAR Inc., 2012; Rodriguez & Green, 1997). The Beck Depression Inventory – Second Edition (BDI-II) measures the presence and intensity of depressive symptoms. Although the constructs of this measure are easily discernable in the content of its items, it may be used in custody evaluations to help evaluators determine the impact of depressive symptoms on childrearing abilities. It contains 21 items that describe various aspects of depressive symptoms across cognitive, affective, and physiological domains, and requires the respondent to describe the frequency and intensity with which these symptoms have occurred in the past two weeks (Jaffe & Mandeleew, 2011).

In addition to parenting self-report measures, evaluations also commonly include interviews and direct observations (Milner & Murphy, 1995). Both structured and unstructured interviews may be used to assess a parent's risk level for physical abuse perpetration by directly asking the alleged perpetrator questions regarding the alleged incident and other risk factors. Additionally, observational methods that are used can be naturalistic, where the parent is directly observed interacting with a child in a natural setting such as his or her home, or structured where a task is set up and the parent's interactions are observed (Milner & Murphy, 1995). Despite their usefulness, the structured and unstructured interviews often employed by evaluators are subject to bias based on the interviewer's beliefs and personal experiences, and the same is true of both devised and naturalistic observational methods (Milner & Murphy, 1995).

Use of personality testing in dependency cases. According to McCann et al. (2008), the typical custody evaluation involves psychological evaluation of each parent/step-parent, and child, as well as observations of interactions between the parental figures and children. Surveys targeting methods employed by child custody evaluators have noted an increase in the use of psychological testing. Researchers have hypothesized that this increase is due to the perception that many parents involved in child custody litigation are psychologically impaired (Stolberg & Kauffman, 2015). The goal of using psychological testing is to gather information that will help evaluators make a recommendation regarding what is in the best interest of the child or children (McCann et al., 2008). Additionally, questionnaires measuring personality and psychological functioning indirectly yield pertinent information regarding a litigant's

parenting ability (Stolberg & Kauffman, 2015). Psychological measures are limited in that those undergoing evaluation are highly motivated to present themselves as functioning better than they are. As such, it is standard to use several different instruments as well as additional information from interviews and outside sources when conducting a child custody evaluation (Stolberg & Kauffman, 2015). The Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) and its revisions has been the most widely-used measure in the evaluation of parents involved in custody disputes and dependency cases. Due to its extensive use, it will be described in more detail in a later section of this paper.

The Millon Clinical Multiaxial Inventory – 3rd Edition (MCMI-III; Millon, Davis, & Millon, 1996) is a self-report clinical inventory that has been met with much controversy throughout the years of its use in child custody/dependency evaluations. Knowledge of the strengths and weaknesses of this measure are paramount to thorough evaluation in custody evaluation settings and in other contexts. For instance, the MCMI-III was specifically designed to measure personality traits. Some evaluators believe its assessment of these factors is more extensive and coherent compared to other measures that also examine personality factors (Plake & Impara, 2001). Additionally, the MCMI-III has exhibited adequate reliability and validity comparable to other effective measures, and has far fewer test items than several other personality measures. The MCMI-III also contains scales assessing for the presence of symptoms consistent with various Diagnostic and Statistical Manual-Fourth Edition (DSM-IV) diagnoses. Despite its connection to various DSM-IV disorders, the MCMI-III is

fundamentally based on the test developer's theory. Evaluators have also found the scoring process to be unduly complicated, which can have an adverse effect on the validity of the test scores (Plake & Impara, 2001). Although the MCMI-III has been used with increasing frequency in recent years, questions regarding its use in custody evaluations arose for the reason that it was normed using only clinical samples and has a tendency to overpathologize individuals not currently receiving psychological treatment (McCann et al., 2008). Historically, it has also been criticized for its questionable validity, which in the past has undermined the credibility of results yielded from its use (McCann et al., 2008). According to the discussion of norms in the manual, the MCMI-III normative sample included those undergoing evaluations for child custody, criminal competency, and personal injury (Millon et al., 1996). Thus, although the normative sample is described as a clinical sample, many researchers maintain that it is suitable for use with the aforementioned populations (Halon, 2001). However, the manual specifies that the MCMI-III should only be used with those for whom psychological difficulties are suspected, or who are currently receiving psychotherapy. From this, it can be concluded that the MCMI-III, when utilized in child custody evaluations, yields the most accurate results when used with clients who are also clinical patients at the time of evaluation (Halon, 2001). Later editions of the manual describe the controversy surrounding use of the MCMI-III in custody evaluations and urge evaluators to examine both perspectives before deciding whether use of the MCMI-III is an appropriate measure to use for their purposes (Millon, Millon, Davis, & Grossman, 2009). Some researchers, including McCann and

colleagues (2008), argue that child custody litigants being evaluated do, in fact, constitute a clinical population because of the focus on determining the presence of psychopathology in parents. This argument is especially relevant in custody evaluations conducted as a result of alleged or substantiated physical abuse, as researchers have often noted physically abusive parents tend to have inadequate abilities to regulate negative emotions and maintain interpersonal relationships (McCann et al., 2008).

Another objective personality measure, the Personality Assessment Inventory (PAI; Morey, 1991), has been used in both criminal and civil cases with increasing frequency over the last several years. This measure is a multi-scale, self-report inventory designed to assess adults' psychological functioning. In addition to validity scales, the PAI contains clinical scales related to various psychological difficulties, as well as subscales that measure diagnostic elements of these difficulties (Spies, Carlson, & Geisinger, 2010). Edens, Cruise, and Buffington-Vollum (2001) described advantages of using the PAI with offender populations. The first advantage is that the estimated reading level required for accurate use of this measure is lower than for many other measures. This is highly beneficial considering many criminal offenders have a lower education level compared to the general population. Additionally, this measure is among the shorter of the personality measures but does not sacrifice psychometric quality for brevity. Furthermore, studies have indicated several of its clinical variables to be particularly useful by contributing to decision-making of forensic factors, for example, aggression, suicide potential, psychosis, and

psychopathy. Mullen and Edens (2008) noted that the PAI is often used in combination with other personality measures with the goal of determining whether or not psychopathology was present. In their study of published court findings, Mullen and Edens (2008) found difficulties with admissibility of the PAI to be a rarity among the cases examined.

The Rorschach is a performance-based measure requiring participants to provide their perceptions of inkblots. It allows examiners to gather behavior observations in a relatively standardized way while simultaneously gathering data about examinees' personality characteristics and thought processes (Calloway, 2008). It is considered performance-based as it involves completion of a problem-solving task, and it is objective in terms of involving standardized coding. A primary advantage of this measure is that it is relatively free from purposeful manipulation of responses by the examinee (Calloway, 2008). When combined with additional assessment data and behavioral observations, the Rorschach can be useful in identification of psychological disturbance and maladjustment (Weiner, 2005). Although researchers agree there is not a single testing profile that separates capable parents from incapable ones, experts in the field have consistently found some characteristics to be indicative of poorer parenting practices. For example, impulsivity, inflexibility, low stress or frustration tolerance, and egocentricity can contribute to such difficulties as a failure to provide children with the nurturance required for healthy psychological development, and these characteristics can be assessed with the Rorschach (Calloway, 2008; Weiner, 2005). Questions about the Rorschach's

reliability and validity have prevented many examiners from incorporating it into their assessment batteries when conducting custody evaluations (Erard, 2005; Erard & Viglione, 2014). However, research has shown that the Rorschach is psychometrically sound and routinely admitted in court proceedings (Erard, 2005; Weiner, 2005).

Overview of the MMPI, MMPI-2, and MMPI-2-RF. The MMPI (Hathaway & McKinley, 1943) and its revised and restructured editions is the most widely used personality test worldwide. It is used in a variety of applications, most prominently in the assessment of psychological maladjustment. Originally published in 1943, the MMPI is a self-report measure assessing personality characteristics and psychopathology. The original MMPI consists of 566 true/false statements presented in a first-person narrative. Examinees' responses load onto 13 standard scales composed of three validity scales and 10 clinical scales that represented the core of the measure. The validity scales included the Lie scale (L), which measures overly positive self-presentation, the Infrequency scale (F), which measures overreporting or over-exaggeration of psychological disturbance and distress, and the Defensiveness scale (K), which measures respondents' outright denial of psychological disturbance and distress. Other validity scales were subsequently added to this measure. Similar to the L scale, the Positive Malingering Scale (Mp) measures denial of faults commonly admitted in the general population and denied when trying to present more positively. The Social Desirability Scale (Sd) is conceptually the opposite of Mp in that while Mp assesses the denial of negative attributes, Sd assesses the attestation of positive attributes. The 10 clinical scales of the MMPI include Scale 1, Hypochondriasis (Hs),

measuring unease about physical symptoms, Scale 2, Depression (D), measuring the presence of depressive symptoms, Scale 3, Hysteria (Hy), assessing the use of denial and repression as defense mechanisms in the face of distress, Scale 4, Psychopathic Deviate (Pd), measuring reactions to conflict, anger, and nonadherence to societal rules, Scale 5, Masculinity/Femininity (MF), assessing adherence to gender norms, Scale 6, Paranoia (Pa), measuring level of suspiciousness toward others, Scale 7, Psychasthenia (Pt), assessing the presence of excessive worries, anxiety, self-doubt, and resulting tension, Scale 8, Schizophrenia (Sc), assessing for the presence of thought-disordered symptoms and personality characteristics common to patients with Schizophrenia, Scale 9, Hypomania (Ma), measuring mood elevation and racing thoughts typical of manic and hypomanic episodes, and Scale 0, Social Introversion (Si), measuring the level of social comfort and desire to engage in social interactions.

Subsequently, a broad range of supplemental scales for the MMPI were developed (Harris & Lingoes, 1955; 1968). The Harris-Lingoes subscales were created to help evaluators better interpret the clinical scales by providing measures of the heterogeneous factors that make up the various scales. These subscales were created for clinical Scales 2, 3, 4, 6, 8, and 9 that had heterogeneous content. Scales 1 and 7 were not given Harris-Lingoes subscales as the creators found these clinical scales to already measure mostly homogeneous factors. Scales 5 and 0 were also not given Harris-Lingoes subscales because they were largely considered to be nonclinical personality scales (Friedman, Bolinskey, Levak, & Nichols, 2014). Over the years the core structure of the MMPI was enhanced by the development of a host of content and

supplementary scales. Wiggins (1966) introduced a set of 13 content scales designed to measure specific areas of dysfunction, such as social maladjustment, phobias, and poor morale. Other research-based supplementary scales were also introduced that have demonstrated psychometric soundness and have remained in use over time. The Welsh Anxiety (Welsh A) and Welsh Repression (Welsh R) scales were developed by Welsh (1956) with the intention of measuring two core underlying dimensions, anxiety and general unhappiness (Welsh A) and internalization and overcontrol of emotion (Welsh R). Another scale that has withstood the test of time is the MacAndrew Alcoholism Scale (MAC; MacAndrew, 1965), measuring the likelihood of substance abuse.

The Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Graham, Ben-Porath, Tellegen, Dahlstrom, & Kaemmer, 1989) was published as a revision of the MMPI. This revision involved updating many of the test questions through improving the wording of numerous items, and developing contemporary norms. Many scales were utilized from the original MMPI, such as the core clinical scales and the Harris-Lingoes subscales; however, new content scales were developed. A revised version of the MAC scale was also included, the MacAndrew Alcoholism Scale – Revised (MAC-R). The MMPI-2 consists of 567 items organized into validity, clinical, content, and supplementary scales. Although the clinical scales remained largely the same, their revision and the addition of new validity scales represented a significant effort by the developers to improve the measure from the previous version.

The MMPI-2 validity scales included the MMPI L, K, and F scales. It also has a newly added F Back (F_B) scale consisting of F scale items on the second half of the test; this section of the test contains face-valid content-driven items that are easier for respondents' to manipulate their answers. Other validity scales designed for the MMPI-2 included the Variable Response Inconsistency Scale (VRIN), which was created to detect inconsistent and contradicting responses, and the True Response Inconsistency Scale (TRIN), designed to detect biased responding in either an acquiescent or nay-saying direction. The Infrequency-Psychopathology Scale, abbreviated F(p), was designed specifically for the MMPI-2 to be used in conjunction with F in order to detect overreporting of psychological distress and disturbance. Previously called the Fake Bad Scale (FBS), the Symptom Validity Scale was also created for the MMPI-2 as another measure to detect overreporting of psychological symptoms. The Superlative Self-Presentation Scale (S), also developed specifically for the MMPI-2, measures positive self-presentation.

The items of the MMPI-2 also load onto nine Restructured Clinical scales, 15 Content scales, 12 Supplementary scales, five Personality Psychopathology Five (PSY-5) scales, and 27 Harris-Lingoes subscales. The Restructured Clinical scales were designed to improve upon the measurement of the core constructs of each clinical scale through removal of items representing shared distress variance. A new set of 15 content scales were designed to measure symptoms such as anxiety, depression, cynicism, anger, and low self-esteem. The Supplementary scale set contains a combination of MMPI scales such as Welsh's A and R and newer measures

such as Addiction Admission (AAS) and Addiction Potential (APS) to assess substance abuse problems. The PSY-5 scales were derived from dimensional characteristics of various psychological disorders and measure aspects of dysfunction along a spectrum rather than presence or absence of these characteristics.

The Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF; Ben-Porath & Tellegen, 2008), is a new, restructured version of the MMPI. This self-report questionnaire consists of the nine Restructured Clinical Scales used with the MMPI-2, nine Validity scales, three Higher-Order scales, 23 Specific Problem scales, two Interest scales, and revised Personality Psychopathology Five (PSY-5-r) scales. The validity scales allow examiners to determine patterns of responding that diminish the representativeness of the results, including random or inconsistent responding (Variable Response Inconsistency, VRIN-r, and True Response Inconsistency, TRIN-r, respectively) overreporting of difficulties (F-r, Fp-r, Fs, FBS-r, and RBS), and underreporting of common faults and psychological disturbance (L-r and K-r, respectively). The Higher Order scales measure three overarching specific areas of dysfunction, including mood and affect (Emotional/Internalizing Dysfunction; EID), disordered thinking (Thought Dysfunction; THD), and under-controlled behavior (Behavioral/Externalizing Dysfunction; BXD). Ben-Porath and Tellegen (2008) examined each of these dimensions in three separate clinical samples and ultimately determined which of the non-overlapping items best measured these dimensions. The Restructured Clinical scales, originally created for the MMPI-2 and included in the MMPI-2-RF, were

developed as a solution to problems caused by the overlap of items among the original clinical scales. These scales were created for Clinical scales 1, 2, 3, 4, 6, 7, 8, and 9 and measure the core constructs of their corresponding MMPI-2 Clinical scales. The demoralization scale (RCd) was added as a measure of general psychological distress and discomfort. The Specific Problems scales examine problems reflective of diagnostic criteria and are divided into four domains consisting of somatic, internalizing, externalizing, and interpersonal. The Somatic/Cognitive scales include five scales purposed to evaluate preoccupation with one's health, the presence of somatic symptoms, and perceived cognitive difficulties. The Internalizing scales were designed to provide more in-depth assessment of difficulties of mood and affect related to elevated scores on the EID Higher-Order scale and RC scales RCd, RC2, and RC7. The Externalizing scales were designed to examine behavioral aspects of BXD and RC scales RC4 and RC9 to determine areas of difficulty leading to elevations on these scales. The Interpersonal scales address various areas of potential interpersonal difficulty. The Specific Problems scales are often used by evaluators to structure their interpretation of the test results. The PSY-5 scales of the MMPI-2-RF are modified versions of those developed for the MMPI-2 assessing psychopathological personality dimensions. The Interest scales allow examiners to consider nonclinical areas that may be gender-related. Derived from MMPI-2 Clinical scale 5, Masculinity-Femininity, the Interest scales reflect Aesthetic-Literary Interests (AES) and Mechanical-Physical Interests (MEC). Descriptions of the MMPI-2-RF scales can be found in Table 1.

Table 1

The MMPI-2-RF Scales

Scale	Description
Validity Scales	
Cannot Say (?)	Blank or double-marked items
Variable Response Inconsistency Scale (VRIN-r)	Inconsistent or random responding
True Response Inconsistency Scale (TRIN-r)	Response bias
Infrequent Responses (F-r)	Endorsement of responses uncommon in general population
Infrequent Psychopathological Responses (Fp-r)	Endorsement of responses uncommon in psychiatric population
Infrequent Somatic Responses (Fs)	Endorsement of somatic complaints uncommon in medical patient populations
Symptom Validity (FBS-r)	Non-credible somatic and cognitive complaints
Response Bias Scale (RBS)	Non-credible memory complaints
Uncommon Virtues (L-r)	Uncommonly claimed moral attributes or activities
Adjustment Validity (K-r)	Abnormally high levels of psychological adjustment
Higher Order (H-O) Scales	
Emotional/Internalizing Dysfunction (EID)	Difficulties with mood and affect
Thought Dysfunction (THD)	Difficulties with disordered thinking
Behavioral/Externalizing Dysfunction (BXD)	Difficulties with under-controlled behavior
Restructured Clinical (RC) Scales	
Demoralization (RCd)	Prevailing unhappiness/dissatisfaction
Somatic Complaints (RC1)	Wide variety of physical health complaints
Low Positive Emotions (RC2)	Vulnerability to depression
Cynicism (RC3)	Believing others are bad/untrustworthy

(cont.)

(Table 1 cont.)

Scale	Description
Ideas of Persecution (RC6)	Believing others pose a threat to self
Dysfunctional Negative Emotions (RC7)	Maladaptive anger, anxiety, and irritability
Aberrant Experiences (RC8)	Unusual sensory experiences and disordered thinking
Hypomanic Activation (RC9)	Over-activation, aggression, impulsivity, grandiosity
Specific Problem (SP) Scales	
Somatic/Cognitive Scales	
Malaise (MLS)	Perception of poor health
Gastrointestinal Complaints (GIC)	Frequent nausea, upset stomach, poor appetite
Head Pain Complaints (HPC)	Head and neck pain
Neurological Complaints (NUC)	Dizziness, weakness, balance problems
Cognitive Complaints (COG)	Problems concentrating and remembering
Internalizing Scales	
Suicidal/Death Ideation (SUI)	Suicidal ideation and recent attempts
Helplessness/Hopelessness (HLP)	Belief that goals are unreachable and problems unsolvable
Self-Doubt (SFD)	Feeling useless, lacking self-confidence
Inefficacy (NFC)	Believing one to be indecisive, ineffective
Stress/Worry (STW)	Focus on disappointments, trouble with stress
Anxiety (AXY)	Pervasive anxiety, frights, frequent nightmares
Anger Proneness (ANP)	Easily angered, impatient
Behavior-Restricting Fears (BRF)	Fears significantly impeding normal activities
Multiple Specific Fears (MSF)	Phobias
Externalizing Scales	
Juvenile Conduct Problems (JCP)	Problems at school and home, stealing
Substance Abuse (SUB)	Current/past misuse of alcohol or drugs
Aggression (AGG)	Physical aggression, violent behavior
Activation (ACT)	Excitable and high-energy

(cont.)

(Table 1 cont.)

Scale	Description
Interpersonal Scales	
Family Problems (FML)	Conflictual family relationships
Interpersonal Passivity (IPP)	Being unassertive and submissive
Social Avoidance (SAV)	Dislike or avoidance of social events
Shyness (SHY)	Feeling uncomfortable and anxious with others
Disaffiliativeness (DSF)	Disliking others and being around them
Interest Scales	
Aesthetic-Literary Interests (AES)	Literature, music, theater interests
Mechanical-Physical Interests (MEC)	Interests in fixing and building things, enjoying the outdoors and sports
Personality Psychopathology Five (PSY-5) Scales	
Aggressiveness-Revised (AGGR-r)	Instrumental aggression
Psychoticism-Revised (PSYC-r)	Disconnection from reality
Disconstraint-Revised (DISC-r)	Under-controlled behavior
Negative Emotionality/Neuroticism-Revised (NEGE-r)	Anger, insecurity, worry, and fear
Introversion/Low Positive Emotionality-Revised (INTR-r)	Social disengagement, anhedonia

Note: Scales and descriptions adapted from Tellegen and Ben-Porath (2011)

Empirical findings on MMPI-related personality characteristics of maltreating parents. Studies examining the personality characteristics of physically abusive parents often include within their samples parents who have perpetrated other types of maltreatment and parents who have perpetrated multiple forms of maltreatment. However, some noteworthy patterns have emerged in the literature. For example, scores on validity and clinical scales did not usually cross into the clinical

range (Ezzo, Pinsoneault, & Evans, 2007; Resendes & Lecci, 2012). The study by Ezzo et al. (2007) comparing a sample of parents who had perpetrated some type of child matreatment with a custody sample composed of divorced and unmarried couples found the maltreatment sample obtained higher scores on MMPI-2 scales 4, 6, 9, and 0, although these scores did not often reach clinical T-score levels of 65 or higher. Scores on these scales within the subclinical (T scores of 60-64) range, however, may still be informative to both researchers and evaluators regarding maltreating parents' cognitions and behaviors. In this study, those in the maltreating sample often elevated scales 4 and 6 to the subclinical level. Scores in this range on scale 4 indicate individuals who are often independent-thinking sensation-seekers, who have difficulties with emotional intimacy, and who under stress tend to become irritable and selectively report to present a more socially acceptable version of themselves. Scores in this range on scale 6 portray an individual who is hypersensitive, tending to take things personally and, when stressed, see others as being purposefully malicious (Ezzo et al., 2007; Friedman, 2015). The notable differences between the maltreatment and nonmaltreatment samples such as that found by Ezzo et al. (2007) have been supported in other similar studies. Resendes and Lecci (2012) found statistically significant differences between the T scores of a parental competency sample and a child custody sample on eight out of ten MMPI-2 clinical scales. The competency sample produced higher scores on scales 1(Hypochondriasis), 2 (Depression), 4 (Psychopathic Deviate), 6 (Paranoia), 7 (Psychasthenia), 8 (Schizophrenia), 9 (Hypomania), and 0 (Social Introversion). Scales 4, 2, 0, and 8

exhibited the largest effect sizes of the clinical scales. That the parental competency group's clinical scale scores were significantly higher indicates members of this sample may have greater potential for parenting and adjustment difficulties compared to the custody sample. Similar results were noted in a study by Stredny, Archer, and Mason (2006) examining personality characteristics shown on the MMPI-2 and MCMI-III. The MMPI-2 scales that on average showed the highest elevations were scales 4 and 6, although both elevations were below clinical level. Scales RC6 and RC3 were also elevated, reflecting feelings of suspiciousness of others and skepticism with regard to their motives. On the MCMI-III the highest elevations were found on personality scales Histrionic, Narcissistic, and Compulsive, however the authors note that these scores were below clinical range (Stredny et al., 2006). MMPI-2-RF profiles produced by child maltreatment samples have demonstrated similar characteristics. Pinsoneault and Ezzo (2012) compared the MMPI-2-RF profiles of parents evaluated in the context of a custody evaluation with the profiles of those evaluated in the course of child maltreatment cases and parental fitness evaluations. Those in the maltreatment and parental fitness group exhibited higher scores on scales RC3 (Cynicism), RC4 (Antisocial Behavior), JCP (Juvenile Conduct Problems), and FML (Family Problems). This group also exhibited higher scores on RC6 (Ideas of Persecution), THD (Thought Dysfunction), RC8 (Aberrant Experiences), and PSYC (Psychoticism). The authors asserted that these findings uphold those of previous research using the MMPI-2 wherein higher elevations were found in child maltreatment groups on scales 4, 6, and 8. This is due to the shared content measured by scales RC3, RC4, JCP, and

FML on the MMPI-2-RF and scales 4, 6, and 8 on the MMPI-2 (Pinsoneault & Ezzo, 2012).

Defensiveness in Personality Test Profiles

Putzke, Williams, Daniel, and Boll (1999) proposed that profiles of all self-report measures fall along a continuum ranging from accurate reporting to either conscious or unconscious attempts to present oneself in the best light. People are sometimes motivated to underreport their symptoms and minimize any appearance of psychological disturbance in order to appear better adjusted and psychologically healthy, a response pattern also known as impression management (Baer, Wetter, Nichols, Greene, & Berry, 1995). A variety of strategies may be employed by respondents attempting to manage the image they present. For example, respondents may deny psychological disturbance or distress, deny common faults, or attribute to themselves characteristics that they believe are seen as more desirable (Arce, Farina, Seijo, & Novo, 2015). It has been suggested that, based on the average validity and clinical scores seen in previous literature of parents undergoing a custody evaluation, this population is generally more defensive than those in the general population (Baer & Sekirnjak, 1997; Posthuma & Harper, 1998). Although it is possible that some parents in this population are genuinely well-adjusted, psychologically healthy individuals, Posthuma and Harper (1998) find this improbable due to the stressful circumstances under which these evaluations occur.

Measures of defensiveness on the MMPI, MMPI-2, and MMPI-2-RF. When assessing a respondent's profile in a high-pressure context such as custody

evaluation, the evaluator must determine whether the profile is an accurate representation of the respondent's psychological adjustment. Defensive responding, whether done consciously through impression management efforts or unconsciously through self-deceptive responding, can distort the accuracy of a profile. Various measures of defensiveness have been developed to assist evaluators in their task of determining the accuracy of the MMPI, MMPI-2, and MMPI-2-RF profile.

Traditional validity scales of the original MMPI that detect defensiveness, also called underreporting, include L, K, and the F-K index (Baer & Sekirnjak, 1997). The L, F, and K scales were developed as measures of defensive responding on the first edition of the MMPI. The L scale measures denial of moral imperfection, while the K scale measures denial of psychopathology. The F-K index utilized the raw-score difference between these scales to detect underreporting in responding. Previous literature on the MMPI has found a tendency for those who underreport their symptoms to differ from those who do not underreport by approximately one standard deviation on scale F and the F-K index (Baer & Miller, 2002; Baer et al., 1995). For scales L and K, honest and defensive respondents differed by slightly less than one standard deviation on average. Large effect differences have also been seen on supplementary validity scales Wsd and Mp (Baer & Miller, 2002). Other validity measures that gauge the evaluee's response style include the Wiener and Harmon Obvious and Subtle scales (O-S), the Lachar and Wrobel Critical Items (CI), and Gough's F-K Dissimulation Index (F-K), all of which identify both defensive responding and overreporting of difficulties (Barthlow, Graham, Ben-Porath,

Tellegen, & McNulty, 2002). The validity scales developed for the original MMPI were later modified for use on the MMPI-2 (Baer & Sekirnjak, 1997). Significant modifications were made to these scales when the MMPI-2 was restandardized, including changes to or deletion of items that contributed to the L, F, and K scales (Bagby, Rogers, Buis, & Kalemba, 1994).

The MMPI and MMPI-2 validity scales have long been scrutinized for their level of predictive ability, with varying results. One such study conducted by Bagby et al. (1997) compared the MMPI-2 profiles of a sample of patients with schizophrenia who were provided instructions to conceal their symptoms with the MMPI-2 profiles of a student sample also given instructions to underreport. The validity indicators assessed revealed differences in their predictive ability across comparison sets, which the authors interpreted as an indication that different validity indicators may be better suited for some assessment situations rather than others. For example, they found the Edwards Social Desirability Scale (Esd) scale, a research-based measure, to be the most effective detector of schizophrenia patients' attempts to minimize their psychopathology. Based on this result, Bagby et al. (1997) asserted that the Esd scale is well-suited to detect psychopathology in situations without formal corroboration of psychopathology, and where there is motivation to underreport psychopathology. Additionally, Wsd and S were determined to be the most effective at distinguishing between honest profiles and faking good profiles of nonclinical respondents (Bagby et al., 1997). A similar study conducted by Bagby et al. (1994) comparing samples of students and psychiatric inpatients given either fake-good or fake-bad instructions

found statistically significant higher scores on L, K, and Mp when instructed to respond defensively compared to a group instructed to exaggerate or fake symptoms, a control group provided standard instructions, and a clinical sample provided standard instructions (Bagby et al., 1994). A study conducted by Baer et al. (1995) using a sample of nonpatient community members noted higher scores across all MMPI-2 validity scales, including standard and supplemental scales, for those who were instructed to fake good compared to those provided with standard instructions. In fact, those in the fake-good group obtained validity scale scores that were approximately two standard deviations higher than those obtained by the standard instruction group. The results suggested the validity scales are fairly equal in their ability to detect underreporting as there were no significant differences between the scales in percentage of participants who were correctly classified as underreporting (Baer et al., 1995). Despite the similarities, Wsd and S demonstrated significant incremental validity compared to L and K. When analyses utilizing all four measures demonstrated no significant improvements compared to those with only Wsd and S, the researchers concluded Wsd and S were the most effective scales at detecting defensive responding. Based on these results, the authors suggested using Wsd and S to supplement validity findings from L and K (Baer et al., 1995). Comparable conclusions were drawn by Posthuma and Harper (1998) in their study of custody litigants and personal injury litigants, resulting in the authors' suggestion to use the standard validity scales as well as the supplemental validity scales in order to obtain the best representation of the litigant's responding. Although some supplemental

scales used on the MMPI-2 have demonstrated higher average sensitivity rates than scale L, L continues to show the highest rate of specificity (Baer & Miller, 2002). Considering the varied success rates of the traditional and supplemental validity scales, Baer and Miller (2002) suggest focusing interpretive attention on scales L and K, as these have been shown to have reasonable ability to detect underreporting and have a wider body of research supporting them (Baer & Miller, 2002). Effect sizes seen in statistical analyses of MMPI-2 defensiveness scales affirm the effectiveness of L and K as detectors of underreporting (Sellbom & Bagby, 2008).

Sellbom and Bagby (2008) illustrated the role of assessment context in underreporting response style through a comparison of two investigations using the MMPI-2-RF. In the first study, a nonclinical sample of university students and a clinical sample with schizophrenia were either instructed to underreport or were given standard instructions. Although L-r and K-r successfully differentiated between underreporters and nonunderreporters on both patient and student samples, and between the underreporting patient sample and standard instruction student sample with large effect sizes, the researchers noted that significantly poorer performance and smaller effect sizes were seen on L-r in differentiating student underreporters from student nonunderreporters. However, this lower effect size was not seen for L-r in study two, which used a student sample provided underreporting instructions within a child custody scenario as well as a differential prevalence sample composed of child custody litigants. The authors suggested the different L-r effect sizes between the studies is indicative of validity indicators being affected by context. Specifically, they

hypothesized that when evaluees attempt to moderate peoples' perceptions of their relationships with others, such as within the context of child custody evaluations, L-r is more impactful (Sellbom & Bagby, 2008).

Methods of Defensiveness Adjustment on the MMPI, MMPI-2, and MMPI-2-RF.

Use of the MMPI K-correction. In addition to developing the validity scales for the purpose of detecting defensive responding, a novel approach was taken by Hathaway and McKinley (1942) in the development of the original MMPI. They created the K scale to be used not only for the detection of defensive responding, but also to adjust for the effects of defensive responding on clinical scales. The goal of the test developers for the K scale was that it would enhance the ability of the clinical scales to identify psychological disturbance while lessening the frequency of profiles falsely appearing within normal limits (Archer, Fontaine, McCrae, 1998). The K scale compensates for defensive responding through the addition of proportions of the K scale raw score total to standard Clinical scales 1, 4, 7, 8, and 9, which are the most susceptible to defensiveness (Archer et al., 1998; Friedman, 2015). The K-correction has been routinely used in MMPI and MMPI-2 profile interpretation.

Silver and Sines (1962) sought to determine whether the use of the K correction or even simply awareness of the K raw score of a profile could aid in producing accurate diagnoses. The researchers utilized the MMPI scores of a sample of patients admitted to a psychiatric hospital and cross-referenced these results with diagnoses given by psychiatric staff prior to examining the patients' results. Raters

received the profiles either with K-corrected scores but no K scale raw score, only the K scale raw score with uncorrected scores, both the K scale raw score and K-corrected scores, or uncorrected K scale scores without the K scale raw score. Raters then sorted the profiles into diagnostic groups based on specified criteria. Under these conditions the researchers did not find that the K scale raw score or the K-corrected scores significantly impacted diagnostic accuracy.

Colby (1989) conducted a study with the purpose of examining whether the original MMPI K correction would affect evaluators' ability to distinguish between the MMPI profiles of patients and nonpatients. The MMPI profiles of a sample of psychiatric inpatients were compared with a sample of Caucasian participants' profiles from a previous study's sample. The results of this study indicated that there were fewer false negatives when distinguishing between patients and nonpatients when the K correction was used, whereas there were fewer false positives when the K correction was not used.

A study by Putzke et al. (1999) assessed the validity of the K-correction in MMPI-2 profiles of end-stage lung disease patients who submitted for evaluation as part of the transplant eligibility determination process. Using a cutoff score of 59 for scale K, the patients were separated into defensive and nondefensive groups. The researchers aimed to determine the ability of the K-correction process to adjust for defensive responding in their sample. When only K-corrected T scores on these scales were used, a significant group difference occurred only on Scale 1, which identifies endorsed physical symptoms, and was noted to be higher for those in the defensive

group. Of the scales that do not receive a K-correction, scale 0 was observed to be significantly lower in the defensive group, while scale 3 was significantly higher. Putzke and colleagues asserted this finding is consistent with previous research findings where defensive respondents attempted to present themselves as more socially outgoing (low scale 0) and affected by physical maladies (high scale 3). The lack of significant differences between the groups on factors external to the testing, including demographic, medical, and psychiatric characteristics, indicates that defensive responding is the primary difference between the groups, as identified by K-corrected scores.

The benefits of the K-correction method has been reconsidered in recent years based on research findings that it introduces some psychometric confounds. For example, McCrae, Costa, Dahlstrom, and Williams (1989) reported that K correction reduced the correlations between scores from MMPI clinical scales and external measures such as NEO Personality Inventory self-report and peer ratings. More recently, Barthlow et al. (2002) argued that the research to date has not shown convincingly that clinical scale scores of clinical samples are more accurate when the K-correction is applied. From a psychometric viewpoint, they asserted that K correction may in fact weaken relationships between clinical scale scores and external criterion measures by removing valid variance from these corrected MMPI scales. Using two samples of outpatient mental health and university clinic clients, they demonstrated that K correction was no better than non-K correction in most cases, and resulted in lower correlations with criterion measures for the subgroup of women

clinic patients. The authors qualified their conclusions with the possibility that K correction may be more useful in circumstances such as child custody evaluations where defensiveness is more prominent, and called for future research to address this issue. Nonetheless, K correction has been discontinued in the MMPI-2-RF.

Derivation of optimal cutting scores. One of the challenges facing evaluators and researchers when attempting to correct for defensiveness is the establishment of optimal cutting scores to achieve this objective. Optimal cutting scores allow for maximum detection of those with psychopathology who are denying their symptoms (i.e. responded defensively) while simultaneously correctly identifying those without psychopathology responding in a manner reflective of their psychological health (Baer & Miller, 2002). This can be quite challenging, as described by Baer and Miller (2002). They noted that, historically, researchers have had difficulty determining cutting scores that are effective across diverse samples. One of the greatest barriers to establishing a universal cutting score is the range of consequences across contexts, including whether an error of false detection or an error of non-detection would be more detrimental to the examinee. According to Baer and Miller (2002), due to this wide range of potential consequences of evaluations, no single set of cutting scores will be optimal across contexts. Support for this suggestion can be found in the study by Putzke et al. (1999), discussed earlier. Although the researchers found defensive responding to be accompanied by lower scores on the clinical scales, they asserted that the standard .5 K-weight they used on the Hypochondriasis scale of the original MMPI

may have been an overcorrection due to the similarities of defensive profiles to profiles of potential transplant recipients who were deemed to be nondefensive.

According to the MMPI-2 technical manual, elevations on scale L in the range of 70 to 79 are somewhat commonplace in clinical settings and do not immediately invalidate the profile, although the profile should be interpreted with caution when the L score is in this range (Butcher et al., 2001). In non-clinical contexts, the validity of the profile is called into question when a T-score over 70 is achieved for this scale. For the K-scale T-scores within the 65 to 74 range are considered questionable, and the same is true for S-scale T-scores in the 70 to 74 range (Butcher et al., 2001; Cooke, 2010). Per the manual's instructions, determining whether the T-scores on scales L, K, and S invalidate the profile partially depends on the elevation of TRIN, and whether the examinee in question is part of a clinical or nonclinical sample (Butcher et al., 2001; Cooke, 2010). With regard to evaluating child custody litigants, previous researchers have proposed a T-score cutoff value of 65 for both L and K in order to account for the higher levels of defensiveness found in this population (Archer et al., 2012).

Researchers continue to strive for the establishment of optimal cutting scores for various MMPI-2 and MMPI-2-RF validity indicators. For example, the study by Bagby Rogers, Buis, and Kalemba (1994), mentioned earlier, compared a sample of outpatients with residual-phase schizophrenia with a sample of college undergraduates. Both groups were provided either fake-good or fake-bad test-taking instructions. They noted that, for scales measuring defensiveness, cutoff scores of O-S

< 18 and F-K < -12 most effectively detected defensive responding. Additionally, as coaching respondents on test-taking strategies has been shown to improve examinees' ability to fool validity indicators, Baer and Miller (2002) found that cutting scores for those who had been coached in successful deception tended to be lower.

Crighton, Marek, Dragon, and Ben-Porath (2017) conducted a study that aimed to establish optimal cutting scores for L-r and K-r that best allow them to predict defensiveness using a simulation design. The researchers utilized MMPI-2-RF data from an archival sample of undergraduate college students and separated them into three groups based on the instructions they were provided by the researchers and their responses to a post-test questionnaire regarding their provided instructions. The three groups were Standard Compliant, who were given standard responding instructions, Underreporting Compliant, who were instructed to respond defensively and did so, and Underreporting Noncompliant, who were instructed to respond defensively and denied having done so on the post-test questionnaire. The Underreporting Compliant group produced significantly lower scores on almost all MMPI-2-RF problem-oriented scales compared to the Underreporting Noncompliant and Standard Compliant groups. Crighton et al. (2017) tested several cutoff scores for L-r and K-r to determine which score produced the fewest false negatives and false positives in their prediction of defensiveness. They found L-r displayed adequate sensitivity of .61 when a cutoff score of 65 was used, and specificity was high across the tested cutoff scores although a cutoff score of 80 produced the highest specificity rate of .99. For K-r, adequate sensitivity at a rate of .54 was found at a cutoff score of 60, and specificity was high

across the examined cutoff scores although a cutoff score of 72 produced a perfect specificity rate of 1.00.

There have been few studies that focus on the development of cutoff scores for the purposes of detection and prediction of defensiveness of the MMPI-2-RF. Instead, studies have typically focused on the other end of the response distortion continuum – detection and prediction of malingering. One example of such a study includes Sellbom and Bagby's (2010) study that compared the MMPI-2-RF responses of undergraduate college students with those of psychiatric inpatients. They focused on scales F-r and Fp-r when developing cutoff scores. Similarly, Rogers, Gillard, Berry, and Granacher (2011) examined MMPI-2-RF responses of a sample of disability referrals using a known-groups design. They, too, sought to establish optimal cutting scores by examining malingering detection scales F-r and Fp-r, as well as scale Fs. A meta-analytic review by Sharf, Rogers, Williams, and Henry (2017) examined the success of studies using the MMPI-2-RF validity scales to predict malingering. They noted the consistent effectiveness of Fp-r in discriminating between malingered and genuine psychopathology, as well as moderate effectiveness of scales FBS-r and Fs.

Empirical findings on MMPI-related defensiveness of maltreating parents. As noted earlier, parents involved in custody evaluations often attempt to minimize or hide psychological distress related to the dissolution of their marriage, such as anxiety and depression. Simultaneously, they may overemphasize positive characteristics such as impulse control. The positive self-presentations may be consciously deceptive or unconsciously biased, which is an important judicial distinction due to the

implications of purposeful deception (Arce et al., 2015). This type of responding may also be common among physically abusive parents, although there is little research specifically targeting this aspect of this particular population. Much of the research including this population is subsumed within the classification of child maltreatment, which can also include child neglect and sexual abuse, and studies regarding parental competency evaluations. However, it is important to take these findings into consideration.

The study conducted by Ezzo et al. (2007) compared MMPI-2 profiles of maltreating parents involved in cases where their parental rights may be terminated with those of non-maltreating parents involved in child custody disputes. Interestingly, while both groups were determined to have responded defensively, those in the maltreatment group obtained scores on scale K that were significantly lower than those of the nonmaltreatment group, while their scores on scale L were significantly higher than the non-maltreatment group with the average score elevated into the clinical range (T=65.94). A later study by Pinsoneault and Ezzo (2012) compared MMPI-2-RF profiles of unmarried and divorced parents with no documented child maltreatment with those of unmarried parents involved in parental fitness evaluations due to documented child maltreatment. Significant differences regarding validity scales were only seen for scale L-r when the maltreatment and nonmaltreatment groups were compared. The average score on scale L-r for the maltreatment group was elevated into the clinical range (T=66.52), whereas the average score on scale L-r for the nonmaltreatment group did not reach even the subclinical level (T=59.32).

Significant differences were not found between the groups on scale K-r; however, given the pattern observed in the study by Ezzo et al. (2007), it is important to note that the nonmaltreatment sample obtained a mean score approximately four points higher than the maltreatment sample on this scale. Additionally, neither group produced scores on this scale that reached the subclinical level.

Similar results were observed in studies with samples of parents court-ordered to undergo psychological evaluations as part of their parental competency determinations. The study by Stredny et al. (2006) examined MMPI-2 and MCMI-III profiles of parents and guardians undergoing psychological evaluations in the course of parental competency determinations. Scores from each measure were compared to their respective normative samples. The average score from the competency sample on scale L of the MMPI-2 was nearly 1.5 standard deviations, i.e. 15 points, above the normative mean and was elevated into the subclinical range (T=64.37). The average score on MMPI-2 scale K from this sample was not elevated and was only 1.5 points above the normative mean (T=51.50). Comparable results were observed in the average MCMI-III profile obtained by this sample as scale Y, which measures socially desirable responding, was elevated. Clinical scale scores from both tests fell below clinical range, with the exception of MMPI-2 scale 4, which was elevated to the subclinical level (T=60.26). An analogous study by Resendes and Lecci (2012) examined parents whose children have been sheltered due to preliminary investigations of "problematic parenting behavior" (p.1055), including allegations of abuse, neglect, unstable mental health, incompetence, and substance abuse. During the

course of these evaluations the sample had been administered the MMPI-2 and provided standard instructions. Their scores were compared with those of a previous study utilizing child custody litigants for whom parental competency is not in question. The average T-score on scale L was elevated into the subclinical range (T=62.6) for the sample of parents under investigation; however, the average T-score for scale K obtained by this same sample was within normal limits. The comparison child custody sample did not display scores on either L or K that reached subclinical or clinical elevations. Notably, the sample of parents under investigation obtained a significantly higher score on scale L compared to that obtained by the sample of custody litigants, while the custody litigant sample obtained a significantly higher score on scale K. Importantly, for the sample of parents under investigation, all clinical scales were below the subclinical and clinical range with the exception of scale 4 which was elevated to the subclinical range (T=63.3).

Siegel, Bow, and Gottlieb (2012) conducted a MMPI-2 study using a sample of parents involved in high-conflict child custody cases court-ordered for psychological evaluation to help determine the best interests of the children. High-conflict custody battles included cases that involved protective orders, protracted litigation, and/or allegations of domestic violence, mental illness, child abuse, or substance abuse. The scores of this sample were compared to the data collected from previous studies using samples of parents involved in custody conflicts. The female subset of the high-conflict sample obtained T scores on scales L, K, and S of approximately 60, which is one standard deviation above the normative sample and significantly higher than

averages found on these scales in previous studies' data sets of custody litigants. The men in the high-conflict sample obtained T-scores on scales L, K, and S significantly higher than the normative and custody litigation samples. Interestingly, their average scores on K and S were higher than those of the women, although their average scores on scale L were approximately two points lower than those of the women. Based on these findings, the authors suggested high-conflict custody litigants unconsciously tended to deny personal faults in addition to consciously attempting to present themselves in the most positive light.

Numerous studies have supported the claim that child custody litigants, even those without allegations of maltreatment, aim to present themselves more positively compared to other groups, most notably seen on scales L, K, and S. These studies have identified two main types of impression management response styles, including defensive response styles and minimizing response styles, both of which lead to underreporting of psychological disturbance and distress (Goldstein & Posthuma, 2015). Additionally, previous research has established that custody litigants tend to elevate validity scales L and K on the MMPI-2 with F-scale T-scores tending to be lower (Archer, Hagan, Mason, Handel, & Archer, 2012). The previously-discussed sample used by Bagby et al. (1999) of custody litigants undergoing psychological evaluation in the process of custody and access hearings exemplified this pattern, although this sample did not include maltreating parents. On average, the sample in Bagby and colleagues' (1999) study produced T-scores within or approaching the subclinical range (approximately one standard deviation above the normative sample)

on scales L and K, respectively. Scores on Scale F, however, were on average below that of the normative mean (T=47.7). Bagby et al. (1999) noted that participants produced at least moderate elevations on both the L and K scales, regardless of whether they were classified as underreporting or non-underreporting. They hypothesized that this finding could speak to a persistent undercurrent of defensiveness within this population (Bagby et al., 1999).

Based on the findings of these studies, there are some similarities between response patterns of maltreating parents, parents whose competency is in question for various reasons including, at times, maltreatment, parents involved in high-conflict custody litigations, and parents involved in custody litigation where evaluation is ordered. The defensive responding seen in these studies suggests that the pressures to appear psychologically healthy and well-adjusted come into play in the context of evaluation. One observed difference between the response patterns of the various sample types is worth noting. Parents involved in cases of documented maltreatment, substance abuse, or questionable psychological functioning responded in a manner so as to present themselves in the best light, thereby elevating scale L; however, they did not respond in a manner reflective of overt denial of psychological maladjustment and only moderately elevated scale K.

Rationale and Hypotheses

Studying the personality and psychological functioning of physically abusive parents can help psychologists identify areas of maladjustment. It can also help identify those parents who are at greater risk for future offending due to problematic psychological deficits and difficulties. However, when a child enters into the dependency system due to physical abuse by his or her parent, and the parent is asked to submit to a psychological evaluation, a significant complication often arises in the parent's response patterns. Parents involved in this process frequently respond defensively, which makes it difficult for evaluators to determine to what degree there is personal maladjustment behind the abusive actions of the parent. When accurate information cannot be gathered regarding the parent's functioning, specifically regarding areas of dysfunction, it is challenging to determine what dysfunction is present that may be contributing to problems in the home, and how best to address this dysfunction and provide support in needed areas. Additionally, it can make it difficult to determine the best placement for the child and whether termination of parental rights is necessary.

MMPI-based studies have noted defensiveness is customary in the responses of maltreating parents as well as nonmaltreating custody litigants. The research examining the degree of defensiveness within these groups using the MMPI-2-RF have generally found maltreating parents to be more defensive than nonmaltreating custody litigants. However, there continues to be a gap in the literature for two primary reasons. The first is that defensiveness in physically abusive parents is not

distinguished from defensiveness in parents who have perpetrated various and sometimes multiple types of maltreatment. Studies specifically examining defensive testing profiles of physically abusive parents have been notably absent from the literature. The second reason is that while most of the extant studies using the MMPI-2-RF have researched the detection of defensiveness using the appropriate validity scales, they have not taken the next step. That is, they have not worked to develop a solution to the complications that arise when trying to interpret defensive MMPI-2-RF profiles. With consideration given to these reasons, the purpose of the current study was twofold. First, the current study aimed to examine the level of defensiveness of a sample of physically abusive parents and compare it to a sample of nonabusive custody litigants to determine whether this sample of physically abusive parents would be more defensive than nonmaltreating custody litigants. Second, the current study intended to establish empirically-derived optimal cutting scores in order to compensate for the defensiveness typically found in the responses of physically abusive parents.

In keeping with the goals of this study, three hypotheses were tested in the first phase of the study:

1. The MMPI-2-RF profiles of the sample of physically abusive parents will be defensive as indicated by L-r and K-r, with defensiveness determined by mean T-scores of 65 or greater for L-r and of 60 or greater for K-r. Scales L-r and K-r were used as they are well-established measures of defensiveness and assess for

uncommonly claimed virtuousness and denial of psychological disturbance, respectively.

2. There will be a suppression effect on the MMPI-2-RF RC scales for the physically abusive parent sample, with scores falling at or below T=55, with the exception of RC6, which was hypothesized to be elevated to at least T=60 based on findings of previous literature.

3. The physically abusive parent sample will produce significantly higher scores on scales L-r and K-r compared to the scores of the comparison sample of custody litigants on scales L-r and K-r. They will also score significantly higher on L-r and K-r than a community sample of parents; this latter comparison was undertaken as a check for the expected pattern of defensiveness.

The central purpose of this study was to attempt to adjust for defensiveness in the MMPI-2-RF Restructured Clinical scale scores of physically abusive parents. This required the development of empirically-derived cutting scores that maximize a differentiation between defensive (physically abusive sample) and nondefensive (Community sample) parents. The goal of establishing optimal cutting scores for the physically abusive parent population was to enable future evaluators to differentiate between the effects of defensive responding on MMPI-2-RF clinical profiles from profiles reflecting nondefensive responding, thus allowing for more accurate interpretation of their test scores.

Methods

Participants

The participants for this study consisted of three samples of parents age 18 and older. The primary sample consisted of biological parents, biologically-related guardians, and non-biological others serving in either a stepparent or guardian role who had been adjudicated to dependency proceedings pursuant to Florida Statute (Chapter 39) due to child physical abuse. MMPI-2-RF based inclusion criteria for this sample consisted of a Cannot Say score of 14 or lower, and VRIN-r and TRIN-r scores of 79 or lower, reflecting the absence of substantial response omissions, inconsistencies, or biases. The 62 participants who met these criteria had a mean age of 31.59 (SD=8.5; range=19-55). The sample was somewhat evenly split between men (43.5%) and women (56.5%). The majority of the sample was Caucasian participants (45.2%), followed by African-American (30.6%), Hispanic (21.3%), and Asian participants (1.6%). Among participants for whom employment information was available (*n*=60), 58.3% were employed and 41.7% were unemployed. Biological parents were the primary child physical abuse perpetrators in this sample, including biological fathers (30.6%) and biological mothers (56.5%). Stepfathers were also among the abuse perpetrators (6.5%), as well as a grandmother, a grandfather, and an uncle (1.5% each).

The vast majority of these physically abusive participants (80.6%) were referred for evaluations in partial fulfillment of their court-ordered reunification plans. Of those who were willing to discuss their history with Child Protective Services

(*n*=58), 60.3% denied and/or had no record of having ever been investigated for any maltreatment allegations, 31.0% had been investigated for maltreatment allegations 1 to 5 times, and 8.6% had been investigated at least 6 times. Although the participants in this sample were adjudicated to dependency proceedings due to allegations of physical abuse, a sizeable proportion of the parents (*n*=30, 48.4%) were unwilling to disclose their involvement in the injuries sustained by their child, despite those injuries having been ruled non-accidental by medical professionals. The remainder of the sample (*n*=32, 51.6%) disclosed a variety of abusive actions, including hitting their child with one or more objects (53.1%), hitting their child with their hand (21.9%), kicking their child (3.1%), shaking their child (9.4%), or another specified action (12.5%).

Table 2 presents additional information about the physical abuse sample in relation to their circumstances at the time of evaluation as well as their personal histories.

Table 2

Physical Abuse Sample Demographics

Demographic Variable	*n*	%
Referral Source		
Family Services	60	96.8
Private Attorney	2	3.2
Number of Children in the Home		
1 to 2	40	64.5
3 or more	21	33.8

(cont.)

(Table 2 cont.)

Demographic Variable	*n*	%
Number of Child Victims		
1	49	79
2 or more	13	21
Victim of Domestic Violence		
Yes	24	38.7
No	37	59.7
Perpetrator of Domestic Violence		
Yes	25	40.3
No	36	58.1
History of Outpatient Treatment		
Yes	19	30.6
No	43	69.4
History of Substance Abuse Treatment		
Yes	4	6.5
No	58	93.5
History of Physical Abuse		
Yes	17	27.4
No	45	72.6
History of Neglect		
Yes	4	6.5
No	58	93.5
History of Child Sexual Abuse		
Yes	8	12.9
No	54	87.1
Witnessed Domestic Violence as a Child		
Yes	9	14.5
No	53	85.5

As seen in Table 2, the majority of the physical abuse sample reported no personal history of physical abuse, sexual abuse, or neglect, nor did they report a history of psychological or substance abuse treatment. Notably, however, a substantial

proportion reported being the victim or perpetrator of domestic violence, 39% and 40%, respectively.

The first comparison sample consisted of 64 parents or guardians who had undergone a custody evaluation in the context of a custody dispute, but who had no history of substantiated child maltreatment allegations. This custody group was comprised of 26 men (40.6%) and 38 women (59.4%) who ranged in age from 23 to 52, with a mean age of 35.88 (SD=6.3). The majority of the parents and guardians in this sample was Caucasian (87.5%), and the remainder was Hispanic (12.5%). African American and Asian persons were not represented in this sample. Approximately half of this sample (51.5%) obtained up to a high school diploma or General Education Diploma (GED), while 45.4% completed at least two years of college. Of those included in this sample, 35.9% had previously been the subject of at least one child custody investigation.

Table 3 provides selected descriptive information for the custody sample alongside that of the physical abuse sample.

Table 3

Key Demographics of the Physical Abuse Sample and Child Custody Comparison Sample

Demographic Variable	Physical Abuse Sample			Child Custody Sample		
	M (SD)	*n*	%	*M (SD)*	*n*	%
Age	31.0 (8.5)	---	---	36.0 (6.3)	---	---

(cont.)

(Table 3 cont.)

Demographic Variable	Physical Abuse Sample			Child Custody Sample		
	M (SD)	*n*	%	*M (SD)*	*n*	%
Gender						
Male	---	27.0	43.5	---	26.0	41.0
Female	---	35.0	56.5	---	38.0	59.0
Ethnicity						
Caucasian	---	28.0	45.0	---	56.0	87.5
Non-Caucasian	---	33.0	53.0	---	8.0	13.0
Education						
High School or Lower	---	34.0	55.0	---	33.0	53.0
Some College or College Degree	---	28.0	45.0	---	29.0	47.0

The second comparison sample consisted of 66 parents or guardians from the community who had custody of their child(ren) and who had not had any complaints or inquiries of child maltreatment. Parents who met exclusion criteria were removed from the final analysis. This consisted of parents whose profiles were excluded due to defensiveness (*n*=2), overreporting (*n*=1), investigation of child maltreatment *(n*=1), and lack of child information (*n*=1). Table 4 presents demographic data for this sample.

Table 4

Key Demographics of the Community Comparison Sample

Demographic Variable	*M (SD)*	*n*	%
Age	38.0 (9.4)	---	---

(cont.)

(Table 4 cont.)

Demographic Variable	*M (SD)*	*n*	%
Gender			
Male	---	16	26
Female	---	45	74
Ethnicity			
Caucasian	---	48	79
Non-Caucasian	---	13	21
Education			
High School or Lower	---	5	8
Some College or College Degree	---	56	92
Age of Child (in months)			
Age Child 1	120.0 (97.4)	---	---
Age Child 2	141.0 (63.4)	---	---
Age Child 3	111.0 (54.3)	---	---
Number of Children in the Home			
1 to 2	---	47	77
3 or more	---	14	22.9
Marital Status			
Single	---	3	5
Cohabitating	---	3	5
Married	---	47	77
Separated or Divorced	---	7	11.5
Widowed	---	1	1.6
Victim of Domestic Violence			
Yes	---	7	12
No	---	53	87
Perpetrator of Domestic Violence			
Yes	---	1	2
No	---	60	98

(cont.)

(Table 4 cont.)

Demographic Variable	*M (SD)*	*n*	%
History of Physical Abuse			
Yes	---	8	13
No	---	53	87
History of Neglect			
Yes	----	3	5
No	---	57	93
History of Child Sexual Abuse			
Yes	---	11	18
No	---	50	82
Sexual Assault in Adulthood			
Yes	---	5	8
No	---	56	82

Inclusion criteria for the two comparison samples based on MMPI-2-RF validity scale guidelines were as follows: Cannot Say <15, VRIN-r and TRIN-r <80, F-r <120, Fp-r <100, Fs<100, FBS-r<100, L-r <80, and K-r<70. These latter criteria regarding the requirements for scales F-r, Fp-r, Fs, FBS, L-r, and K-r were applied to these samples to ensure that their profiles were not invalidated by either overreporting or underreporting of psychological difficulties.

Instruments

MMPI-2-RF (Tellegen & Ben-Porath, 2008)

The primary measure used in this study was the MMPI-2-RF. The technical manual for this measure outlines the process used to derive its sound psychometric properties. As the 338 items on the MMPI-2-RF were taken from the original 567 items from the MMPI-2, test developers were able to utilize data collected from the MMPI-2 normative sample as well as an outpatient community mental health sample,

a psychiatric inpatient sample, and male psychiatric inpatients from a Veteran Administration's hospital. For the MMPI-2-RF Validity scales, the test-retest reliability coefficients ranged from .40 for scale TRIN-r to .84 for scale K-r. The Standard Error of Measurements (SEMs) for the Validity scales on the normative sample ranged from 4 on scales F-r and K-r to 8 on scale TRIN-r. Internal consistency reliability for the Validity scales from the normative sample displayed different ranges for both men and women. For men, internal consistency reliability ranged from .37 on scale TRIN-r to .69 on scale F-r. For women, internal consistency reliability ranged from .20 on scale VRIN-r to .71 on scale F-r. For the Higher-Order (H-O) scales, test retest reliability ranged from .71 on scale THD to .91 on scale BXD, with the SEMs ranging from 3 on scales EID and BXD to 5 on scale THD. Internal consistency reliability for the H-O scales on the normative sample ranged from .69 on scale THD to .88 on scale EID for men, and from .82 on scale BXD to .94 on scale EID for women. Test-retest reliability for the Restructured Clinical (RC) scales ranged from .64 on scale RC6 to .89 on scale RC4, with SEMs ranging from 3 on scale RC4 to 6 on scale RC6. Internal consistency reliability on the RC scales for men ranged from .63 on scale RC6 to .87 on scale RCd, and for women ranged from .63 on scale RC2 to .89 on scale RCd. For the Somatic/Cognitive scales, test-retest reliability ranged from .54 on scale NUC to .82 on scale MLS, with SEMs ranging from 4 on scale MLS to 7 on scale NUC. For men internal consistency reliability ranged from .52 on scale NUC to .64 on scales GIC and COG, and for women ranged from .58 on scale NUC to .69 on scales GIC and COG. The Internalizing scales displayed test-retest reliability ranging

from .65 on scale HLP to .85 on scale MSF, with SEMs ranging from 4 on scales SFD, NFC, ANP, and MSF to 6 on scales SUI, HLP, and BRF. Internal consistency reliability for men ranged from .39 on scale HLP to .72 on scale ANP, and for women ranged from .34 on scale SUI to .73 on scale NFC. Overall, these coefficients reflect acceptable reliability levels, with the caveat that lower values are expected for scales with fewer items.

Data documenting the external validity of the MMPI-2-RF, as reported in the technical manual, were gathered from a variety of settings where the MMPI-2-RF was likely to be used, including clinical, forensic, medical, and non-clinical settings. Utilizing data from external sources, including therapist ratings, other objective self-report measures, and record reviews, Tellegen and Ben-Porath (2008) demonstrated appropriate convergent and discriminant validity. Examination of the convergent validity of the MMPI-2-RF with the MMPI-2 showed expected associations between the MMPI-2 scales and the MMPI-2-RF scales, without so much similarity that the restructuring could be seen as redundant. Overall, the psychometric properties of the MMPI-2-RF indicate this measure reliably and validly measures response styles, personality characteristics, and psychopathology of respondents.

Procedure

Following approval from the Institutional Review Board of Florida Institute of Technology and the Doctoral Research Project Committee, demographics and test results for the primary sample were collected from archival data from an outpatient forensic psychology practice with which the chair of this study maintains a research

relationship. Practitioners at that site frequently conduct evaluations using the MMPI-2 and, more recently, the MMPI-2-RF in the context of physical abuse, sexual abuse, neglect, and child custody evaluations. The data were collected solely from records of parents in the process of completing Reunification Case Plans following loss of custody of their child(ren) due to commission of physical abuse. As informed consent was obtained at the time of evaluation, all parents in the aforementioned group were eligible for inclusion. For members of this sample who were evaluated utilizing the MMPI-2, their evaluations were re-scored to the MMPI-2-RF. The scored responses were entered into a confidential Statistical Package for the Social Sciences (SPSS) database for analysis. Participants were assigned a numerical code in place of identifying information to adhere to guidelines set by the Health Insurance Portability and Accountability Act (HIPAA) of 1996.

The first comparison sample for this study consisted of a subset of an archival database comprised of parents who participated in a child custody evaluation, but who had not lost custody of their children or been accused of physical abuse. Demographic and test score data were previously collected and entered into a confidential SPSS database developed by the chair of this Doctoral Research Project. As informed consent had been obtained at the time of evaluation, all members of this sample were eligible for inclusion in this study. The subset of this sample was derived by first eliminating all parents who did not meet inclusion criteria based on the MMPI-2-RF validity scale guidelines. Then, frequency data for key demographic features of the Physical Abuse sample were used to guide the derivation of the subset of cases from

the Child Custody sample. This process ensured that the samples were reasonably similar in basic demographic makeup.

A second comparison sample was collected by this student researcher and consisted of a community sample of 61 parents in Brevard County, Florida who had no record with DCF for maltreatment allegations. This sample was recruited by initially posting flyers in local community centers, including the Wickham Park Community Center, the Lipscomb Community Center, the Joseph N. Davis Community Center, the Palm Bay Community Center. Additionally, postings on the Thrifty Moms of Brevard social media page (*n*=2), posting of flyers on the Florida Institute of Technology Campus (*n*=2), handing out of flyers at children's sports camps (*n*=8), and word of mouth (*n*=54) yielded the desired sample size. The flyers included information about the study, the researcher's contact information, and information about an incentive for participants. Informed consent was obtained from each participant prior to their inclusion in the study. Participants were administered the MMPI-2-RF in small group testing sessions that were held at public libraries, in research-purposed rooms on Florida Institute of Technology's campus, or at another community activity location. Their responses were then computer scored. To ensure accuracy, two participants' responses were selected for hand scoring to verify that computer scoring yielded identical results. The data collected from this sample was also entered into the SPSS database for analysis. To maintain the anonymity of participants, each person was identified solely by a number, and informed consents were separated from testing and demographic data.

Data Analyses

Preliminary analyses consisted of computing descriptive statistics including means, standard deviations, and percentage data to describe the characteristics of each sample. Additionally, the means and standard deviations of MMPI-2-RF scores were computed separately for each of the three samples. MMPI-2-RF T-score means for the physically abusive parent sample were used to address Hypotheses 1 and 2. A Multivariate Analysis of Variance (MANOVAs) followed by univariate Analyses of Variance (ANOVA) were conducted to compare L-r and K-r scores of the Physical Abuse sample and the Custody sample, and the Physical Abuse sample and the Community sample.

A second part of the statistical analysis involved submitting MMPI-2-RF test scores of the Physical Abuse and Community samples into Receiver Operating Characteristics (ROC) analyses to determine optimal cutoff scores for the MMPI-2-RF RC scales for the physical abuse group. The ROC analysis allows for estimates of test-score sensitivity and specificity. Sensitivity rates were calculated to determine the probability that a scale score above the cutoff score correctly identified the presence of psychological difficulties. Specificity rates were calculated to determine the probability that a scale score below the cutoff score correctly identified the absence of psychological difficulties. The Area Under the Curve (AUC) was utilized to evaluate the classification accuracy of the MMPI-2-RF for those selected cutoff scores, where higher values for the AUC indicated a higher likelihood of correct classification as clinically elevated or non-elevated. Following the guidelines of Streiner and Cairney

(2007), an AUC value of .70 was set as the minimum acceptable value, as significant AUC values of 1 indicate perfect accuracy. A power analysis was conducted to determine the minimum number of participants required for an ROC analysis. When power (1-β) was set at 0.80 and α=0.05, the analysis indicated at least 24 participants in each group, for a total of 48 participants, is required to prevent false acceptance of the null hypothesis. The current samples consisted of *N*=62 in the Physical Abuse group and *N*=61 in the Community group. Thus, the ROC analysis had sufficient statistical power to produce effective differentiation between the two groups.

Results

The initial step in data analysis was to calculate the test score means and standard deviations for the primary sample. Table 5 presents the descriptive statistics regarding MMPI-2-RF scores of the Physical Abuse total sample, and separately for men and women in the sample.

Table 5

Physical Abuse Sample MMPI-2-RF T-score Means and Standard Deviations

	Total Sample N=62		Men n=27		Women n=35	
Scale	Mean	SD	Mean	SD	Mean	SD
Validity Scales						
VRIN-r	50.5	10.9	50.9	9.1	50.2	12.3
TRIN-r	57.1	8.4	55.1	5.3	58.6	10.0
F-r	52.4	18.8	48.0	13.0	55.8	21.9
Fp-r	51.9	13.4	49.8	10.3	53.6	15.3
Fs	53.2	17.5	48.2	8.4	56.9	21.5
FBS-r	56.3	13.4	50.0	6.2	61.2	15.4
L-r	70.6	14.0	70.7	13.3	70.5	14.8
K-r	57.1	10.8	57.7	9.8	56.6	11.6

(cont.)

(Table 5 cont.)

	Total N=62		Men n=27		Women n=35	
Scale	Mean	SD	Mean	SD	Mean	SD
Higher-Order Scales						
EID	45.9	9.9	44.0	6.1	47.3	11.9
THD	51.6	13.3	50.6	10.5	52.3	15.3
BXD	46.6	8.5	49.0	9.5	44.7	7.3
Restructured Clinical Scales						
RCd	47.2	10.4	45.0	5.5	48.8	12.9
RC1	57.2	12.3	55.3	9.2	58.7	14.2
RC2	47.2	9.8	46.7	7.3	47.5	11.5
RC3	50.2	11.9	51.3	11.2	49.3	12.5
RC4	49.2	10.2	51.1	11.5	47.7	9.0
RC6	56.2	14.3	56.8	12.4	55.8	15.8
RC7	45.8	12.6	43.4	9.6	47.7	14.3
RC8	49.4	12.7	47.4	9.2	50.9	14.8
RC9	43.3	9.2	43.9	8.1	42.8	10.0
Somatic/ Cognitive Scales						
MLS	49.3	9.0	48.4	6.3	49.9	10.7
GIC	53.3	12.8	50.3	8.9	55.7	14.8
HPC	53.9	12.1	53.2	10.2	54.3	13.5
NUC	53.9	13.6	51.6	9.5	55.7	16.0
COG	48.5	12.0	45.4	8.2	50.8	14.0
Internalizing Scales						
SUI	46.2	7.4	45.0	0.0	47.2	9.9
HLP	45.4	8.4	44.6	6.9	46.0	9.5
SFD	47.0	8.2	46.0	7.6	47.9	8.6
NFC	46.8	10.8	43.7	7.9	49.1	12.1
STW	47.6	10.8	46.8	9.8	48.2	11.6
AXY	50.7	12.9	47.0	8.2	53.5	15.1
ANP	45.4	9.8	44.8	7.3	45.9	11.4
BRF	52.1	13.2	48.6	8.0	54.7	15.7
MSF	50.1	9.1	45.6	6.3	53.5	9.4

(cont.)

(Table 5 cont.)

	Total N=62		Men n=27		Women n=35	
Scale	Mean	SD	Mean	SD	Mean	SD
Externalizing Scales						
JCP	51.8	10.4	54.7	10.5	49.5	10.0
SUB	45.3	6.9	46.2	8.0	44.6	5.9
AGG	44.0	9.0	44.2	10.0	43.8	8.3
ACT	44.8	12.6	45.3	8.5	44.4	15.1
Interpersonal Scales						
FML	46.8	10.8	44.2	8.1	48.8	12.3
IPP	46.6	8.7	44.7	8.2	48.0	8.9
SAV	51.0	9.9	50.6	8.0	51.3	11.2
SHY	44.6	8.6	42.7	7.2	46.1	9.3
DSF	49.7	8.9	47.0	6.6	51.7	9.9
Interest Scales						
AES	46.7	8.5	43.2	8.5	49.4	7.6
MEC	51.4	10.2	58.4	9.0	46.0	7.5
Personality Psychopathology Scales						
AGGR-r	52.7	8.9	56.2	10.2	50.1	6.9
PSYC-r	49.1	13.8	47.7	10.8	50.2	15.7
DISC-r	47.7	9.5	52.3	10.8	44.2	6.5
NEGE-r	47.1	11.3	45.9	8.1	47.9	13.3
INTR-r	50.6	8.4	50.6	6.8	50.6	9.6

For the substantive scales, which include Higher Order, Restructured Clinical, Somatic/Cognitive, Internalizing, Externalizing, Interpersonal, Interest, and Personality Psychopathology scales, and excluding the Interest scales, T-score means ranged from 44 to 57 for the total sample, which was within one standard deviation from the mean. The range was from 42 to 56 for men, and from 42 to 58 for women. Thus, this sample's mean scores were largely congruent with the norms of the MMPI-2-RF.

To test Hypothesis 1, means and standard deviations were examined for the L-r and K-r scores of the primary sample. Table 5 includes these values, shown in the Validity Scales cluster. Examination of the MMPI-2-RF profiles for the primary sample of physically abusive parents revealed that the mean T-score for L-r met the expected defensiveness criteria of $\geq$ 65 T (M = 70.6, SD = 14.0). Further examination of the MMPI-2-RF profiles for this sample revealed that the mean T-score for K-r did not reach the expected cutoff score of 60 (M = 57.1, SD = 10.8). However, it was half a standard deviation above the normative mean, thus placing it within the moderately defensive range. Hence, Hypothesis 1 was largely confirmed.

Hypothesis 2 was tested through analysis of means and standard deviations of RC scale scores for the physical abuse sample. Table 5 again includes these values. It was expected that the means for all RC scale scores, with the exception of RC6, would fall below a T-score of 55. Results showed this for the majority of the RC scale scores, as mean scores were in the range of 43.3 (SD = 9.2) to 50.2 (SD = 11.9). However, the T-score for RC1 was above the expected cutoff of 55 (M = 57.2, SD = 12.3). Although the RC6 mean score was half of a standard deviation above the normative mean (M = 56.2, SD = 14.3), it did not reach the expected cutoff score of $\geq$ 60. Thus, Hypothesis 2 was also largely, but not completely, confirmed.

Hypothesis 3 was evaluated through a MANOVA conducted on mean scores for the dependent variables L-r and K-r. The independent variable was the parenting group into which the participants were classified. After examining essential properties of the data set, it was found that the required statistical assumptions were satisfied,

including continuous dependent variables, distinct categorical groups within the independent variable, independence of observations, adequate sample size, lack of outliers, multivariate normality, linear relationship between variables, homogeneity of variance-covariance matrices, and lack of multicollinearity. The MANOVA result was significant, (Wilk's $\lambda = .564$, $F[4, 366] = 30.38$, $p < .001$, partial $\eta^2 = .249$), indicating that group membership significantly impacted defensiveness. Univariate analyses of variance (ANOVAs) revealed parenting group significantly affected L-r scores, $F(2, 184) = 47.70$, $p < .001$, partial $\eta^2 = .341$. Tukey HSD post-hoc tests showed that parents in the Physical Abuse group ($M = 70.6$, $SD = 14.0$) displayed significantly higher T-scores on L-r than the Custody comparison sample ($M = 55.3$, $SD = 9.7$) and the Community comparison sample ($M = 51.6$, $SD = 10.2$). Additionally, ANOVAs revealed parenting group significantly affected K-r scores, $F(2, 184) = 15.30$, $p < .001$, partial $\eta^2 = .143$. Tukey HSD post-hoc tests indicated that parents in the Physical Abuse group ($M = 70.6$, $SD = 14.0$) obtained significantly higher T-scores on K-r than the Community comparison sample ($M = 51.6$, $SD = 10.2$). There was no significant difference between the Physical Abuse group and Custody comparison sample on K-r T-scores. In light of the significant differences observed between the Physical Abuse, Custody, and Community groups, Hypothesis 3 was largely confirmed.

The primary analysis of the data was conducted using Receiver Operating Characteristics (ROC) analyses. Sensitivity and specificity were evaluated for various T-scores to determine the optimal cutting score for each RC scale. Analyses were repeated using MMPI-2-RF RC scale raw scores for verification purposes. Ultimately

T-scores were selected for analysis as they are used when interpreting MMPI-2-RF profiles. Table 6 presents AUC values and effect sizes represented by Cohen's *d*, alongside means and standard deviations for RC scales for the Physical Abuse and Community samples.

Table 6
Comparison of MMPI-2-RF RC Scales for Physical Abuse and Community Samples

			Physical Abuse Sample N=62		Community Sample N=61	
RC Scale	AUC	*d*	Mean	SD	Mean	SD
RCd	.705	-0.618	47.2	10.4	53.6	10.5
RC1	.616	-0.298	57.2	12.3	60.3	7.5
RC2	.641	-0.431	47.2	9.8	51.4	10.1
RC3	.479	0.130	50.2	11.9	48.9	7.7
RC4	.593	-0.294	49.2	10.2	52.2	9.9
RC6	.451	0.275	56.2	14.3	51.4	14.5
RC7	.702	-0.560	45.8	12.6	52.5	11.4
RC8	.596	-0.201	49.4	12.7	51.7	10.7
RC9	.564	-0.278	43.3	9.2	45.7	8.4

The AUC values for scales RCd at 0.705 (standard error [SE] = .0.48, $p < .001$, 95% confidence interval [CI]: 0.611-0.798) and RC7 at 0.702 (SE = 0.048 , $p < .001$, 95% CI: 0.609-0.795) were observed to meet the minimum acceptable threshold at AUC $\geq$ .70. A one-way ANOVA revealed a significant between-group difference for scale RCd, $F(1, 121) = 11.7$, $p = .001$, with a medium effect size, $d = -0.618$. A significant between-group difference was also found for scale RC7, $F(1,121) = 9.6$, $p < .05$,and produced a medium effect size, $d = -0.560$. No significant between-group differences were produced by the other RC scales. Additionally, AUC values for the remaining scales were below the acceptable AUC value, and the between-group

differences on these scales produced small effect sizes. Scales RC3 and RC6 produced particularly low AUC values and performed below chance values.

Table 7 illustrates the cut scores obtained through the ROC analysis for each RC scale that represented the best balance between sensitivity and specificity. Sensitivity was prioritized when balanced values could not be obtained due to the need for accurate identification of those in the Physical Abuse group. The values provided in this table are based on participants who scored at and below the targeted cutting score.

Table 7

Optimal Cutting Scores for MMPI-2-RF RC Scales for the Physical Abuse Sample

RC Scale	Optimal Cut Score	Hit Rate	Sensitivity	Specificity	PPP	NPP
RCd	50	.675	.710	.639	.710	.639
RC1	59	.585	.581	.607	.710	.459
RC2	48	.626	.694	.574	.694	.557
RC3	50	.463	.516	.410	.516	.410
RC4	51	.545	.565	.525	.565	.525
RC6	50	.463	.419	.508	.419	.508
RC7	47	.634	.645	.623	.645	.623
RC8	50	.561	.581	.541	.581	.541
RC9	47	.520	.710	.328	.710	.328

As shown in Table 7, optimal cutting scores for the RC scales were largely close to the MMPI-2-RF normative mean of 50. Scales RC2, RC7, and RC9 had optimal cutting scores slightly below the normative mean. The sensitivity for RC2 at a cutoff score of 48 was 69.4%, and the specificity was 57.4%. Scales RC7 (sensitivity 64.5%, specificity 62.3%) and RC9 (sensitivity 71.0%, specificity 32.8%) both had optimal cutting scores of 47. Indeed, RC1 was the only scale for which the optimal cutting score nearly reached the higher subclinical range (55-59) for this measure. At

an optimal cutting score of 59, the sensitivity for RC1was 58.1% and the specificity was 60.7%. Scales RCd and RC7 tended to have the most balanced sensitivity and specificity values, while scales RC3, RC6 and RC9 produced the most imbalanced and/or insufficient values with either sensitivity or specificity below chance levels.

Using the optimal cut scores found through the ROC analysis as a point of reference, statistical qualities of potential alternative cut scores were explored using cross tabulation analyses, shown in Table 8. From the specific cut points, hit rate was calculated and sensitivity, specificity, positive predictive power (PPP), and negative predictive power (NPP) were determined.

Table 8

Hit Rate, Sensitivity, Specificity, Positive Predictive Power (PPP), and Negative Predictive Power (NPP) for RC Scale Cutoff Scores

RC Scale	Cut Score	Hit Rate	Sensitivity	Specificity	PPP	NPP
RCd	<45	.683	.780	.634	.561	.852
	<50	.675	.667	.684	.710	.639
	<55	.577	.554	.645	.823	.328
RC1	<55	.585	.622	.564	.452	.721
	<60	.585	.571	.609	.710	.459
RC2	<45	.585	.628	.563	.435	.738
	<50	.634	.623	.648	.694	.574
RC3	<45	.480	.476	.481	.323	.639
	<50	.463	.471	.455	.516	.410
	<55	.455	.473	.406	.694	.213
RC4	<45	.561	.618	.539	.339	.787
	<50	.545	.547	.542	.565	.525
	<55	.561	.545	.600	.774	.344

(cont.)

(Table 8 cont.)

RC Scale	Cut Score	Hit Rate	Sensitivity	Specificity	PPP	NPP
RC6	<45	.463	.464	.463	.419	.508
	<50	.463	.464	.463	.419	.508
	<55	.463	.464	.463	.419	.508
RC7	<45	.650	.667	.636	.613	.689
	<50	.699	.671	.740	.790	.607
RC8	<45	.602	.667	.571	.419	.787
	<50	.561	.563	.559	.581	.541
	<55	.561	.547	.595	.758	.361
RC9	<45	.512	.517	.508	.484	.541
	<50	.528	.521	.556	.806	.246

As seen in the table above, scales RCd, RC2, and RC7 produced the highest hit rates compared to those of the other RC scales. These scales also obtained the highest combined sensitivity and specificity of the specified cut scores. Interestingly, at a cut score of 50, RC7 obtained a higher sensitivity, 67.1%, and specificity, 74.0%, than the originally obtained cut score of 47, which obtained a sensitivity of 64.5% and a specificity of 62.3% (shown in Table 7).

Discussion

Physically abusive parents are rarely studied as a single population outside the context of a maltreatment sample. As such, distinguishing personality characteristics of this population from those of maltreating parents as a whole can be challenging. This is problematic when conducting research-informed evaluations, particularly when considering that previous research has shown physically abusive parents themselves to be a heterogeneous group (Francis et al., 1992). Additionally, physically abusive parents often respond defensively on measures of personality assessment, which can render their results unusable to evaluators (Ezzo et al., 2007; Pinsoneault & Ezzo, 2012; Resendes & Lecci, 2012; Stredny et al., 2006). Despite the research that has been conducted illustrating defensive responding in this population, methods to correct for the effects of the defensiveness have not been explored. Hence, the goal of this study was to compare a sample of physically abusive parents with a control sample of nonabusive parents from the community, and subsequently establish optimal cutting scores to be used with defensive MMPI-2-RF profiles of physically abusive parents.

The first hypothesis of this study centered on examining the expectation that the Physical Abuse group would be defensive. Defensive responding within this group was analyzed through mean T-scores on validity scales L-r and K-r. The scores of these scales were hypothesized to reach a certain criterion level based on findings from the previous literature. Studies comparing various child maltreatment samples indicated mean L-r scores would be elevated into the clinical range, while K-r scores would be elevated into the subclinical range (Ezzo et al., 2007; Pinsoneault & Ezzo,

2012; Stredny et al., 2006). Results showed mean L-r T-scores for the physically abusive parent sample in the current study exceeded the expected cutoff score. Mean K-r T-scores did not meet the expected cutoff score; however, it was elevated to half a standard deviation above the mean, thus indicating a moderate level of defensiveness. The scores obtained by this sample are similar to scores produced by similar samples in the previous research literature. Studies that examined MMPI-2 and MMPI-2-RF scores of maltreating parents who were evaluated in similar contexts also tended to produce L and L-r scores that reached clinical significance, while K and K-r scores tended to fall within normal limits (Ezzo et al., 2007; Pinsoneault & Ezzo, 2012). This pattern of responding indicates parents in the current Physical Abuse sample tended to respond defensively due to denial of personal faults and demonstrated an overall positive self-presentation, rather than denial of psychological symptoms or maladjustment. The observed socially desirable responding as measured by L-r may not be purposeful, but rather an unconscious bias reflective of poor insight and a low level of self-awareness (Arce et al., 2015, Friedman, 2015). Conversely, this pattern of responding may be reflective of purposeful intentions to deny even commonplace shortcomings given the context of the psychological evaluations. These respondents may be concerned about potentially negative legal ramifications that may extend to compromised custody arrangements or reunification plans in relation to admission of their flaws. High elevations on L-r, such as those produced by the current sample, are uncommon among the general population. Evaluators who encounter this pattern of responding may expect to see rigidity as well as simultaneous hypervigilance to

conformity (Friedman et al., 2015). The mean K-r score of the present sample suggests respondents generally felt in control of their lives and their reactions to their circumstances. However, research has previously suggested that K-r scores within the observed range may be more indicative of defensiveness when obtained by those from lower socioeconomic statuses. This may be due to the higher prevalence of stressors that can lead to maladjustment (Friedman et al., 2015). However, data on the socioeconomic status of respondents was not available for this sample. Regardless, it is unlikely that the relatively lower K-r scores of the physical abuse sample indicate healthy psychological adjustment, given the circumstances under which these parenting evaluations take place (Posthuma & Harper, 1998)

The second hypothesis predicted a suppression effect would be evident in the RC scale scores of the Physical Abuse sample. This hypothesis was informed by previous studies with similar samples in which the clinical or RC scales were rarely elevated even into the subclinical range. The exception to this was RC6, or clinical scale 6 in studies using the MMPI-2, in which this scale was often elevated within the subclinical range. Largely, the RC scores of the Physical Abuse sample adhered to the expectations of this hypothesis in terms of falling below a T-score of 55. The exception was RC1, which was elevated above the anticipated level, nearly reaching the subclinical range. This was unexpected given the results of previous studies that typically found the suppression effect impacted all clinical or RC scales when defensiveness was present (Baer & Sekirnjak, 1997; Ezzo et al., 2007; Pinsoneault & Ezzo, 2012; Stredny et al., 2006). However, it fit the findings of Putzke et al. (1999),

discussed earlier. Considering the constructs measured in RC, it is not unthinkable that this score should be somewhat high within this population. Elevation on RC1 indicates the presence of somatic discomforts typically reflective of psychological distress (Friedman, 2015). It is commonly acknowledged within the medical and psychological communities that stress can manifest as physical symptoms (Mayo Clinic, 2016). Additionally, it is often easier, both socially and psychologically, for people to acknowledge and report their difficulties in the form of physical symptoms rather than as symptoms of psychological maladjustment. Considering the elevation of RC1 despite evident defensiveness, it is reasonable to conclude that the scores obtained by the Physical Abuse sample are indicative of felt distress and difficulty.

The third hypothesis of this study addressed the expectation that the Physical Abuse sample would exhibit greater defensiveness in responding compared to the Custody sample and the Community sample. Validity scales L-r and K-r were used in this comparison. Indeed, the Physical Abuse sample produced a significantly higher mean score on L-r compared to the Custody sample. This result indicated that impression management efforts by the Physical Abuse sample are even higher than the Custody comparison sample. It is reasonable to expect some level of defensiveness secondary to the high stakes of the evaluative circumstances under which both of these samples were tested. However, in addition to risks regarding determinations of custody and allowable contact with the examinee's child(ren) that were present for both samples, risks to those in the Physical Abuse sample also included potential legal consequences, particularly if the parents' abusive actions toward their children

represent substantial risk of child injury or even death. The lack of significant difference between the Physical Abuse and Custody groups on scale K-r indicates that impression management through denial of personal faults, rather than denial of psychological maladjustment, was the primary response strategy for both groups. This may speak to the specific risks found within the evaluation contexts, as discussed previously. This pattern of responding is particularly interesting given the somewhat elevated RC1 score produced by the Physical Abuse group. In fact, this finding lends credence to the idea that the abusive respondents find it acceptable to express their psychological difficulty somatically. Additionally, the Physical Abuse group produced significantly higher scores on scales L-r and K-r compared to the Community sample. This result was expected and allowed for further comparison between the samples. The result obtained in this study was similar to the findings of the study by Pinsoneault and Ezzo (2012) that compared maltreatment and nonmaltreatment samples and found the maltreatment sample produced higher scores on both scales. Notably, the between-group differences on K-r were statistically significant in the current study, although they were not in Pinsoneault and Ezzo's (2012) study.

Following the evidence of significant between-group differences in defensive responding between the Physical Abuse and Community samples, the RC scores obtained by each group were compared. Significant differences between the groups were only found on scales RCd and RC7, both of which measure psychological distress. A notable suppression effect was observed through comparison of the mean scores obtained on these scales, in that the Physical Abuse group had significantly

lower scores on RCd and RC7 compared to the Community group. Considering the stressful circumstances under which respondents in the Physical Abuse group were evaluated, and the contrasting, relatively stress-free testing circumstances of the Community group, the observed differences should have logically been in the opposite direction. Given the previously observed suppression effect on the RC scores of the Physical Abuse group as predicted in Hypothesis 2, this result was expected, albeit seemingly paradoxical.

The most important results of this study pertained to the ROC analyses conducted with the purpose of determining optimal cutting scores. The ultimate goal is for the cutting scores to be used in future evaluations of physically abusive parents in an effort to adjust for defensive responding in interpretations. This analysis revealed scales RCd and RC7 had the largest AUC values of the RC scales at .705 and .702, respectively. Indeed, RCd and RC7 were the only scales for which the AUC value reached the minimum threshold of 0.70, as recommended in the literature (Streiner & Cairney, 2007). These findings paralleled those from the between-group analyses that demonstrated the effectiveness of RCd and RC7 in distinguishing between the Physical Abuse and Community groups. This result indicates that the most efficacious differentiation between the groups was obtained when the abusive sample produced lower scores than the nonabusive sample on the emotional indicator scales. This lower level of reactivity might at first glance seem paradoxical, given the gravity and potential implications of the charges against the physically abusive parents. However, when considering the defensiveness evident in the profiles produced by the Physical

Abuse sample, this not surprising. Other AUC values ranged from .564 to .641, except for RC3 and RC6, for which AUC values were below chance values, therefore reflecting that these indices of cynicism and ideas of persecution are not particularly discriminatory between abusive and nonabusive samples.

Optimal cutting scores for the RC scales ranged between 47 and 50, except for RC1, which had an optimal cutting score of 59. For RCd and RC7, they were 50 and 47, respectively. These cutting scores for RC scales are considerably lower than the typical cutting scores of $T \geq 60$ and $T \geq 65$ used in clinical profile analysis of MMPI-type measures. In fact, except for RC1, the optimal cutting scores for the RC scales are at or below the level of the normative mean. The fact that these cutting scores are lower reflects the impact of defensive responding on the physically abusive parents' scores. Ultimately, it would behoove future evaluators to be aware of the RC scale score suppression among physically abusive parents and use these cutting scores to correct for their defensiveness, particularly for scales RCd and RC7. The more a scale is suppressed by defensiveness, the lower the cutting score should be in order to accurately reflect the level of psychological disturbance and maladjustment, as suggested by Archer, Hagan, Mason, Handel, and Archer (2012) in their study comparing the MMPI-2 and MMPI-2-RF. Therefore, given the level of suppression noted on scales RCd and RC7, lower cutting scores are essential to improving the accuracy of interpretation.

It may also be beneficial to establish a threshold score for scales L-r and K-r, as was done for scales L and K in the Archer et al. (2012) study, to determine when

the lower cutting scores for the RC scales would be appropriate. This was not undertaken in the current study as setting lower thresholds for L-r and K-r complicates interpretation of the MMPI-2-RF profile. Specifically, it would likely result in many more test profiles being of dubious validity or not interpretable, which would lower the utilization rate of the test overall. The alternative, then, becomes that of using other data and non-test clinical impressions, which are limited in reliability and validity. Therefore, a better compromise for the purposes of this study was to set the optimal cutting scores for substantive scales, such as the RC scales. Thus, the test profile is rendered interpretable to some degree in evaluations of physically abusive parents.

The current study has several positive features. To begin with, it offers a profile of MMPI-2-RF scores of physically abusive parents, which can serve as reference data both in future investigations of this population as well as for evaluators to consult when conducting evaluations. It is also the first of its kind to compare a specific physical abuse sample to a custody sample, rather than comparing a custody sample to a heterogeneous sample that includes respondents who have perpetrated various forms of child maltreatment. The primary contribution of this unique pairing is that it sets the stage for future research into this dyad using alternative measures that further explore their similarities and differences. Future research should be undertaken to explore specifically the role of denial in these populations, and its implications. Additionally, future research could further explore the impact of various interventions employed with this population, and whether insight-oriented techniques reduce or strengthen use of impression management in responses.

One limitation of this study was its sample size and geographic specificity in which the data was collected. Given the observed heterogeneity of physically abusive parents as a group (Francis et al., 1992), diversification and enlargement of the sample may strengthen or alter the results found here. Another limitation is the lack of prior research to consult that used a sample solely comprised of physically abusive parents. Further research should be conducted to verify or disconfirm the results found here. Overall, the current study contributes to the developing literature on identifying and adapting for test defensiveness when evaluating physically abusive parents with the MMPI-2-RF.

References

American Psychological Association (2010). Guidelines for child custody evaluations in family law proceedings. *American Psychologist, 65*(9), 863-867.

American Society for the Positive Care of Children. (2016). Child abuse statistics and facts in the U.S.. Retrieved from https://americanspcc.org/child-abuse-statistics/

Arce, R., Farina, F., Seijo, D., Novo, M. (2015). Assessing impression management with the MMPI-2 in child custody litigation. *Assessment, 22*(6), 769-777.

Archer, R.P., Fontaine, J., McCrae, R.R. (1998). Effects of two MMPI-2 validity scales on basic scale relations to external criteria. *Journal of Personality Assessment, 70*(1), 87-102.

Archer, E.M., Hagan, L.D., Mason, J., Handel., & Archer, R.P. (2012). MMPI-2-RF characteristics of custody evaluation litigants. *Assessment, 19*(1), 14-20.

Baer, R.A. & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2. *Psychological Assessment, 14*(1), 16-26.

Baer, R.A. & Sekirnjak, G. (1997). Detection of underreporting on the MMPI-2 in a clinical population: Effects of information about validity scales. *Journal of Personality Assessment, 69*(3), 555-567.

Baer, R.A., Wetter, M.W., Nichols, D.S., Greene, R., Berry, D.T. (1995). Sensitivity of MMPI-2 validity scales to underreporting of symptoms. *Psychological Assessment, 7*(4), 419-423.

Bagby, R.M., Nicholson, R.A., Buis, T., & Radovanovic, H. (1999). Defensive responding on the MMPI-2 in family custody and access evaluations. *Psychological Assessment*, *11*(1), 24-28.

Bagby, R.M., Rogers, R., Buis, T., & Kalemba, V. (1994). Malingered and defensive response styles on the MMPI-2: An examination of validity scales. *Assessment, 1*(1), 31-38.

Bagby, R.M., Rogers, R., Nicholson, R.A., Buis, T., Seerman, M.V., & Rector, N.A. (1997). Effectiveness of the MMPI-2 validity indicators in the detection of defensive responding in clinical and nonclinical samples. *Psychological Assessment, 9*(4), 406-413.

Barthlow, D.L., Graham, J.R., Ben-Porath, Y.S., Tellegen, A., & McNulty, J.L. (2002). The appropriateness of the MMPI-2 K-correction. *Assessment, 9*(3), 219-229.

Ben-Porath, Y. S., & Tellegen, A. (2008). MMPI-2 Restructured Form: Manual for administration, scoring, and interpretation. Minneapolis, MN: University of Minnesota Press.

Berkout, O.V., Kolko, D.J. (2016). Understanding child directed caregiver aggression: An examination of characteristics and predictors associated with perpetration. *Child Abuse & Neglect, 56*, 44-53.

Black, D.A., Heyman, R.E., & Slep, A.M. (2001). Risk factors for child physical abuse. *Aggression and Violent Behavior, 6*, 121-188.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B.(2001). *MMPI-2: Manual for administration and scoring (Revised Edition)*. Minneapolis, MN: University of Minnesota Press.

Calloway, G.C. (2008). The Rorschach: Its use in child custody evaluations. *Journal of Child Custody, 2*(1), 143-157.

Child Abuse and Neglect Reporting Act, 1 U.S.C. §§ 11165.3-11174.3 (1987).

Choca, J.P. & Widiger, T.A. (2001). Test review of the Millon Clinical Multiaxial Inventory – III. In B.S. Plake & J.C. Impara (Eds.), The fourteenth mental measurements yearbook [electronic version]. Retrieved from the Burros Institute's Mental Measurements Yearbook online database.

Colby, F. (1989). Usefulness of the K correction in MMPI profiles of patients and nonpatients. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 1*(2), 142-145.

Cooke, G. (2010). MMPI-2 defensiveness in child custody evaluations: The role of education and socioeconomic level. *American Journal of Forensic Psychology, 28*(2), 5-16.

Crighton, A.H., Marek, R.J., Dragon, W.R., & Ben-Porath, Y.S. (2017). Utility of the MMPI-2-RF validity scales in detection of simulated underreporting: Implications of incorporating a manipulation check. *Assessment, 24*(7), 853-864.

Damashek, A., Nelson, M.M., & Bonner, B.L. (2013). Fatal child maltreatment: Characteristics of deaths from physical abuse versus neglect. *Child Abuse & Neglect, 37*, 735-744.

Department of Children and Families (2012). *Child welfare pre-service training participant guide*. Tampa, FL.

Edens, J.F., Cruise, K.R., Buffington-Vollum, J.K. (2001). Forensic and correctional applications of the Personality Assessment Inventory. *Behavioral Sciences and the Law, 19*(4), 519-543.

Erard, R.E. (2005). What the Rorschach can contribute to child custody and parenting time evaluations. *Journal of Child Custody, 2*(1), 119-142.

Erard, R.E. & Viglione, D.J. (2014). The Rorschach Performance Assessment System (R-PAS) in child custody evaluations. *Journal of Child Custody, 11*(3), 159-180.

Ezzo, F.R., Pinsoneault, T.B., & Evans, T.M. (2007). A comparison of MMPI-2 profiles between child maltreatment cases and two types of custody cases. *Journal of Forensic Psychology Practice, 7*(2), 29-43.

Francis, C.R., Hughes, H.M., & Hitz, L. (1992). Physically abusive parents and the 16-PF: A preliminary psychological typology. *Child Abuse & Neglect, 16*, 673-691.

Friedman, A. F., Bolinskey, P. K., Levak, R. W., & Nichols, D. S. (2015). Psychological assessment with the MMPI-2/MMPI-2-RF, 3rd ed. New York: Routledge/Taylor & Francis Group.

Goldstein, M.L. & Posthuma, A. (2015). Use of the Paulhus Deception Scales and MMPI-2 with custody litigants. *American Journal of Forensic Psychology, 33*(3), 19-34.

Halon, R., (2001). The Millon Clinical Multiaxial Inventory-III: The normal quartet in child custody cases. *American Journal of Forensic Psychology, 15*(1), 57-75.

Harris, R. E., & Lingoes, J. C. (1955). Subscales for the Minnesota Multiphasic Personality Inventory: An aid to profile interpretation. (Mimeographed materials.) San Francisco, CA: University of California, Langley Porter Neuropsychiatric Institute.

Harris, R. E., & Lingoes, J. C. (1968). Subscales for the Minnesota Multiphasic Personality Inventory: An aid to profile interpretation (Rev. ed.). (Mimeographed materials.) San Francisco, CA: University of California, Langley Porter Neuropsychiatric Institute.

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory*. Minneapolis: University of Minnesota Press.

Heinze, M. C., & Grisso, T. (1996). Review of instruments assessing parenting competencies used in child custody evaluations. *Behavioral Sciences & the Law, 14*(3), 293-313. doi:3.0.CO;2-0" TARGET="_blank">http://dx.doi.org.portal.lib.fit.edu/10.1002/(SICI)1099-0798(199622)14:3<293::AID-BSL241>3.0.CO;2-0

Herrmann, C.K. & Martin, M.J. (1988). Factors associated with hospitalization in confirmed cases of physical child abuse. *Early Child Development and Care, 31*, 35-41.

Jaffe, A.M. & Mandeleew, D. (2011). Essentials of a forensic child custody evaluation. *Law Trends & News, 7*(3), Retrieved from http://www.americanbar.org/content/newsletter/publications/law_trends_news_practice_area_e_newsletter_home/2011_spring/forensic_custody_evaluation.html

Klevens, J. & Leeb, R.T. (2010). Child maltreatment fatalities in children under 5: Findings from the National Violence Death Reporting System. *Child Abuse & Neglect, 34*, 262-266.

MacAndrew, C. (1965). The differentiation of male alcoholic outpatients from nonalcoholic psychiatric patients by means of the MMPI. *Quarterly Journal of Studies on Alcohol, 26*, 238–246.

Mart, E.G. (2003). The concise approach to child custody assessment. *The Journal of Psychiatry & Law, 31*, 461-478.

Mayo Clinic. (2016, April 28). Stress symptoms: Effects on your body and behavior. Retrieved from https://www.mayoclinic.org/healthy-lifestyle/stress-management/in-depth/stress-symptoms/art-20050987

McCann, J.T., Flens, J.R., Campagna, V., Collman, P., Lazzaro, T., Connor, E. (2008). The MCMI-III in child custody evaluations: A normative study. *Journal of Forensic Psychology Practice, 1*(2), 27-44.

McCrae, R.R., Costa, P., Dahlstrom, W.G., Barefoot, J.C., Siegler, I.C., & Williams, R.B. (1989). A caution on the use of the MMPI K-correction in research on psychosomatic medicine. *Psychosomatic Medicine, 51*, 58-65.

Millon, T., Davis, R. D., & Millon, C. (1996). The Millon Clinical Multiaxial Inventory–III manual. Minneapolis, MN: National Computer System.

Millon, T., Millon, C., Davis, R., & Grossman, S. (2009). MCMI-III: Millon Clinical Multiaxial Inventory-III manual. Bloomington, MN. NCS Pearson, Inc.

Milner, J.S. (1991). Physical child abuse: Perpetrator screening and evaluation. *Criminal Justice and Behavior, 18*(1), 47-63.

Milner, J.S. (1996). Physical child abuse assessment: Perpetrator evaluation. *Criminal Behavior and Mental Health, 6*(2), 191-193.

Milner, J.S. & Chilamkurti, C. (1991). Physical child abuse perpetrator characteristics: A review of the literature. *Journal of Interpersonal Violence, 6*(3), 345-366.

Milner, J.S. & Murphy, W.D. (1995). Assessment of child physical and sexual abuse offenders. *Family Relations, 44*(4), 478-488.

Morey, L. C. (1991). Personality Assessment Inventory professional manual. Odessa, FL: Psychological Assessment Resources.

Mullen, K.L. & Edens, J.F. (2008). A case law survey of the Personality Assessment Inventory: Examining its role in civil and criminal trials. *Journal of Personality Assessment, 90*(3), 300-303.

New York Penal Law, O U.S.C. §§ 260.00-260.34 (2015).

PAR Inc. (2012). Parenting Stress Index - Fourth Edition (PSI-4). Retrieved from http://www4.parinc.com/Products/Product.aspx?ProductID=PSI-4

PAR Inc. (2012). State-Trait Anger Expression Inventory-2 (STAXI-2). Retrieved from http://www4.parinc.com/Products/Product.aspx?ProductID=PSI-4

Perry, M.A., Wells, E.A., & Doran, L.D. (1983). Parent characteristics in abusing and nonabusing families. *Journal of Clinical Child Psychology, 12*(3), 329-336.

Pinsoneault, T.B. & Ezzo, F.R. (2012). A comparison of MMPI-2-RF profiles between child maltreatment and non-maltreatment custody cases. *Journal of Forensic Psychology Practice, 12*, 227-237.

Posthuma, A.B. & Harper, J.F. (1998). Comparison of MMPI-2 responses of child custody and personal injury litigants. *Professional Psychology: Research and Practice 29*(5), 437-443.

Proceedings Related to Children, 5 U.S.C. §§ 39.01-39.15 (2017).

Putzke, J.D., Williams, M.A., Daniel, F.J., & Boll, T.J. (1999). The utility of K-correction to adjust for a defensive response set on the MMPI. *Assessment 6*(1), 61-70.

Resendes, J. & Lecci, L. (2012). Comparing the MMPI-2 Scale scores of parents involved in parental competency and child custody assessments. *Psychological Assessment, 24*(4), 1054-1059.

Rodriguez, C. M., & Green, A. J. (1997). Parenting stress and anger expression as predictors of child abuse potential. *Child Abuse and Neglect, 21*(4), 367-377. Retrieved from https://search-proquest-com.portal.lib.fit.edu/docview/15977985?accountid=27313

Rogers, R., Gillard, N.D., Berry, D.T., & Granacher, R.P. (2011). Effectiveness of the MMPI-2-RF validity scales for feigned mental disorders and cognitive impairment: A known-groups study. *Journal of Psychopathology and Behavioral Assessment, 33*, 355-367.

Schnitzer, P.G., & Weigman, B.G. (2005). Child deaths resulting from inflicted injuries: Household risk factors and perpetrator characteristics. *Pediatrics, 116*, e687-e693.

Sharf, A.J., Rogers, R., Williams, M.M., & Henry, S.A. (2017). The effectiveness of the MMPI-2-RF in detecting feigned mental disorders and cognitive deficits: A meta-analysis. *Journal of Psychopathology and Behavioral Assessment.* doi: 10.1007/s10862-017-9590-1

Sellbom, M. & Bagby, R.M. (2008). Validity of the MMPI-2-RF L-r and K-r scales in detecting underreporting in clinical and nonclinical samples. *Psychological Assessment, 20*(4), 370-376.

Sellbom, M. & Bagby, R.M. (2010). Detection of overreported psychopathology with the MMPI-2-RF [corrected] validity scales. *Psychological Assessment, 22*(4), 757-767.

Siegel, J.C., Bow, J.N., & Gottlieb, M.C. (2012). The MMPI-2 in high conflict custody cases. *American Journal of Forensic Psychology, 30*(3), 21-34.

Silver, R.J. & Sines, L.K. (1962). Diagnostic efficiency of the MMPI with and without the K correction. *Journal of Clinical Psychology, 18*(3), 312-314.

Springer, K.W., Sheridan, J., Kuo, D., & Carnes, M. (2007). Long-term physical and mental health consequences of childhood physical abuse: Results from a large population-based sample of men and women. *Child Abuse and Neglect, 31*(5), 517-530.

Stith, S.M., Liu, T., Davies, C., Boykin, E.L., Alder, M.C., Harris, J.M., Som, A., McPherson, M., & Dees, J. (2009). Risk factors in child maltreatment: A meta-analytic review of the literature. *Aggression and Violent Behavior, 14*, 13-29.

Stolberg, R. & Kauffman, C. (2015) The utility of the MCMI-III in child custody evaluations: The impact of clinical norms. *Journal of Child Custody, 12,* 325-336.

Stredny, R.V., Archer, R.P., & Mason, J.A. (2006). MMPI-2 and MCMI-III characteristics of parental competency examinees. *Journal of Personality Assessment, 87*(1), 113-115.

Streiner, D.L & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry, 52*(2), 121-128.

Texas Family Code, 5 U.S.C. § 261.001 (2005)

Thorpe, G.L., Burrows, R.F., & Cox, A.A. (2010). Test review of the Personality Assessment Inventory. In R.A. Spies, J.F. Carlson, & K.F. Geisinger (Eds.), *The eighteenth mental measurements yearbook* [electronic version]. Retrieved from the Burros Institute's Mental Measurements Yearbook online database.

U.S. Department of Health and Human Services, Administration on Children, Youth, and Families. (2007). *Child maltreatment 2005*.

Weiner, I.B. (2005). Rorschach assessment in child custody cases. *Journal of Child Custody, 2*(3), 99-119.

Welsh, G. S. (1956). Factor dimensions A and R. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine* (pp. 264–281). Minneapolis, MN: University of Minnesota Press.

Whipple, E.E. & Webster-Stratton, C. (1991). The role of parental stress in physically abusive families. *Child Abuse & Neglect, 15*, 279-291.

Wiggins, J. S. (1966). Substantive dimensions of self-report in the MMPI item pool. *Psychological Monographs, 80*(22, Whole No. 630).