9-6-2024

Comparing Caregiver-Teacher-Adolescent Mental Health Reporting: Examining Changes Over Time and Related Outcomes Among Youth Involved in the Child Welfare System

Jennifer Osborne
Loyola University of Chicago Graduate School

LOYOLA UNIVERSITY CHICAGO

COMPARING CAREGIVER-TEACHER-ADOLESCENT MENTAL HEALTH REPORTING: EXAMINING CHANGES OVER TIME AND RELATED OUTCOMES AMONG YOUTH INVOLVED IN THE CHILD WELFARE SYSTEM

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

PROGRAM IN CLINICAL PSYCHOLOGY

BY

JENNIFER L. OSBORNE

CHICAGO, IL

MAY 2024

ACKNOWLEDGMENTS

First, I would like to thank my family. My parents have taken many roles throughout my life, including teachers, cheerleaders, counselors, mentors, and friends. Regardless of their role, they, along with my brother, have always cared for me beyond my accomplishments, which has given me the freedom to take risks and grow into who I am today. My Grandpa Worley was nicknamed “Stats,” as he subscribed to several academic journals for “fun” throughout his life. He, along with my Grandma Worley, modeled lifelong learning. In addition, my Grandpa Osborne taught me to prioritize my education. He, along with my Grandma Osborne, made that education possible. They laid the foundation for my academic journey.

Next, I would like to thank my community. Specifically, the Cubehort, Kate McFeeley, Sam Perry, Amanda Ruggieri, Nicole Hodgkinson, and countless others carried me through this process. I will always be grateful for the countless coffee dates, dance parties, text messages, and silly memes that allowed me to stay grounded, confident, and productive on days where this project felt impossible.

Finally, I would like to thank the Promoting Adjustment in Children through Evaluation (PACE) Lab. Specifically, Dr. Scott Leon and Dr. Nathan Lutz were invaluable throughout this process. I am particularly grateful for Dr. Leon’s support in deciphering the conceptual and the “why” in this project and for Dr. Lutz in guiding me through new statistical terrain and for problem solving R code via email at all hours of the day. Thank you also to my brilliant

committee, Dr. Zoe Smith, Dr. Grayson Holmbeck, and Dr. Tracy DeHart, for their guidance throughout this process.

Without my generous and faithful support system, none of this would be possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The benefits of psychological assessment are particularly salient for children and youth in the child welfare system, as youth in contact with the system are more likely to experience mental health symptoms, disorders, and comorbidities as compared to youth in the general population. These assessments often utilize multiple reporters, as emerging evidence indicates that utilizing multiple more effectively captures the mental health experiences of children and youth. However, the extant literature has consistently observed only low-to-moderate (i.e., .2-to-.3) agreement across informants, including among dyads in contact with the child welfare system. Several factors are associated with level of agreement across reporters, including race, gender, age, behaviors and symptoms of interest, and relationship of the respondent to the child. Specific factors associated with child welfare involvement also contribute to discrepancies across reports, including maltreatment history and foster caregiver/biological parent status. The current study seeks to expand on this research about multi-informant reporter discrepancies by examining reporting discrepancies among caregivers, teachers, and 12-year-old adolescent dyads with prior and/or ongoing child welfare involvement, as well as capturing how teacher-caregiver discrepancies change over time via growth mixture modeling and how trajectory predicts child-rated mental health at age 18.

In keeping with the extant literature, low-to-moderate discrepancies were observed across caregiver-child, child-teacher, and teacher-caregiver dyads regardless of symptom clusters. Discrepancies did not differ according to race or gender. Next, growth mixture modeling

indicated that the best fit for internalizing symptoms was a three-class solution; trajectories included caregiver-teacher dyads with 1) consistently low levels of disagreement, 2) consistently high levels of disagreement, and 3) dyads with high to moderate decreasing disagreement. Regarding externalizing behaviors, the three-class solution was also most appropriate and included caregiver-teacher dyads with 1) consistently low levels of disagreement, 2) low to moderate increasing disagreement, and 3) moderate to low decreasing disagreement. Finally, in terms of total symptoms, the four-class solution was most appropriate and included caregiver-teacher dyads with 1) moderate to high increasing disagreement, 2) low to moderate increasing disagreement, 3) consistently low disagreement, and 3) high to moderate decreasing disagreement, though these trajectories differed depending on whether caregiver or teacher reported higher symptoms. Notably, only trajectories with increasing levels of disagreement predicted child-rated mental health outcomes at 18 when controlling for baseline mental health symptoms. Specifically, when trajectories including increasing discrepancies ending in the moderate range, children reported higher mental health symptomology at age 18; when trajectories including increasing disagreement ending in the high range, children reported significantly lower mental health symptoms at 18. The implications of these findings for research and clinical use are discussed.

CHAPTER ONE

INTRODUCTION

Clinical diagnostic tests have a wide variety of applications in forensic, medical, and psychological settings (Butcher, 2006) and can be used for several purposes across the treatment continuum. First, psychological assessments are frequently used as screening measures as they provide a quick and accurate method for assessing individual mental health functioning (Hunsley & Mash, 2005). Similarly, psychological assessments provide diagnostic clarity, which is important for both diagnosis and treatment planning (Hunsley & Mash, 2020). Finally, using psychometrically sound tests can improve treatment monitoring and evaluation (Mash & Hunsley, 2005), as well as treatment outcomes (Jenson-Doss et al., 2018). Taken together, psychological testing and assessment can elucidate mental health symptoms and guide interventions.

The benefits of psychological testing are particularly salient for children and youth in contact with the child welfare system. As a result of prior and ongoing complex trauma experiences and systemic injustice, youth involved in the child welfare system are more likely to experience mental health symptoms, disorders, and comorbidities than those in the general population (Vasileva & Peterman, 2018). In order to capture the complex mental health symptomology present in this population, the child welfare system typically uses both broadband and narrow assessment instruments to capture mental health functioning children in care. These assessments often utilize multiple reporters, as emerging evidence indicates that prioritizing assessment (i.e., multiple measures) over individual psychological tests more effectively captures the mental

health experiences of children and youth (De Los Reyes et al., 2019). However, despite evidence that multi-informant reporters increase nuance in diagnosis, there is little guidance about how to conceptualize and address disagreement between reporters when it does occur.

The extant literature has consistently observed low-to-moderate (i.e., .2-to-.3) agreement across informants (De Los Reyes, 2015), including among dyads in contact with the child welfare system (Makol et al., 2020). Several factors are associated with level of agreement across reporters, including the child's gender (De Los Reyes & Kazdin, 2005; McWey et al., 2018), age (Smith, 2007), behaviors and symptoms of interest (Hoffman & Chu, 2015), and relationship of the respondent to the child (Achenbach et al., 2019). Specific factors associated with child welfare involvement also contribute to discrepancies across reports, including maltreatment history (McWey et al., 2018) and foster caregiver/biological parent status (McWey et al., 2015).

Discrepancies across reporters often yield meaningful diagnostic data (Strijker et al., 2011). First, disagreement among reporters often capture how specific symptoms vary across settings (Achenbach et al., 2019), which may elucidate specific mechanisms contributing to the mental health symptomology. Additionally, multi-informant assessment discrepancies may capture relationship differences across reporters (van Dulmen et al., 2011), including the relative relational closeness of reporters. This is a particularly important factor to note among youth involved in the child welfare system, as they may help to inform placement decisions and recommended services.

Multi-informant reporting discrepancies are also related to mental health challenges above and beyond the negative outcomes associated with adolescent-onset psychiatric disorders

(Keller et al., 2010). Specifically, the extant literature indicates that Caregiver-child disagreement in reporting is associated not only with continued symptoms, but with increases in symptoms over time (Ferdinand et al., 2004), including among child welfare populations (McWey et al., 2014). Further, disagreement among adolescents and parents influences the course of treatment (De Los Reyes, 2010). Caregiver-child discrepancy is associated with poorer treatment outcomes, particularly when the caregiver reports higher symptomology regardless of symptom type (Goolsby et al., 2018). Relatedly, among child-teacher dyads, there is evidence to suggest that child-teacher reporting discrepancies are associated with fewer treatment improvements among youth with ADHD diagnoses (Hennig et al., 2017). While the nascent literature indicates that multi-informant reporter discrepancies have profound effects on long-term child mental health functioning, much is left to discover as it relates to teacher-caregiver disagreement.

The author is unaware of any study that examines discrepancies in caregiver-child, child-teacher, and teacher-caregiver reporting, how teacher-caregiver discrepancies change over time, and how discrepancy trajectories contribute to specific mental health outcomes in the context of youth with prior child welfare involvement. The dearth of literature related to multi-informant trajectories and outcomes is particularly problematic, as there is evidence to suggest that discrepancies do not remain stable over time (Becker-Haimes et al., 2018).

Thus, the current study seeks to expand on this research about multi-informant reporter discrepancies. First, it quantifies 12-year-old caregiver-child, child-teacher, and caregiver-teacher dyads ratings of internalizing symptoms and externalizing behaviors in the child welfare population as captured by the Child Behavior Checklist (CBCL), the Youth Self-Report (YSR),

and the Teacher Report Form (TRF). Next, discrepancies between caregivers and teachers were also assessed at ages 10, 14, and 16. Using the level of disagreement at each time point, growth mixture modeling was employed to identity latent trajectories existing in the study. Finally, the current study examined how the multi-informant discrepancy trajectories predict child ratings of mental health at 18 years old.

CHAPTER TWO

LITERATURE REVIEW

Psychological Testing: Historical and Current Uses

While psychological testing has been a component of the clinical psychology since its inception in the early 20th century, it was firmly cemented as an essential part of the field following the expansion of psychological services after World War II (Butcher, 2006). Psychological tests are distinct from psychological assessments. Specifically, psychological assessments integrate many pieces of information across domains; psychological tests may provide pieces of that information (American Psychological Association, 2013). However, the field frequently refers to psychological tests as assessments. Though funding for psychological test development has waxed and waned in subsequent decades, mental health services and insurance companies have increasingly included measures in their coverage.

Currently, psychological testing refers to a broad range of measures, most typically presented via standardized questionnaires or interviews. Psychological tests are frequently categorized as either tests of typical behavior (i.e., designed to capture daily functioning) or tests of maximal performance (i.e., designed to capture optimal functioning). Psychological tests are further dichotomized as structured and unstructured, wherein structured tests most commonly use prescriptive response options (e.g., questionnaires) and unstructured tests allow for individuals to provide their own responses (e.g., Rorschach; Medicine et al., 2015). Measures of mental health symptoms are most commonly presented as structured measures of typical behavior.

Clinical diagnostic tests have a wide variety of applications in forensic, medical, and psychological settings (Butcher, 2006). Psychological tests are popular, in part, as a result of their strong, demonstrated psychometric properties and applicability across populations (Bagby & Solomon-Krakus, 2019). For a psychological test to be widely circulated, the test must demonstrate strong reliability, as demonstrated by test-retest consistency, inter-rater reliability, and internal consistency (Medicine et al., 2015). Once distributed, psychological tests are frequently used to (a) screen for possible mental health needs; (b) aid in formal diagnosis, (c) plan treatment, (d) and monitor and evaluate treatment efficacy (Hunsley & Mash, 2005).

Screening measures are often briefer than full psychological tests. However, in order to be effective, screening measures must be sensitive enough to detect individuals who have a psychological disorder and specific enough to differentiate those who do and do not meet clinical criteria (Sheldrick et al., 2015). Screening measures are often implemented during routinized check-ins, such as those included in medical, child welfare, school, and juvenile justice casework practice. These check-ins can be used as opportunities to screen for mental health symptomology among individuals not already connected to mental health services.

Additionally, psychological tests can provide diagnostic clarity. Specifically, tests are particularly effective at assisting in the diagnosis of mental health disorders as they are frequently problem-specific and therefore able to assess specific symptom clusters associated with disorders and help to rule out others (Hunsley & Mash, 2020). In this way, psychological tests can be useful in teasing apart complex, overlapping symptomology and providing diagnostic clarity. For example, difficulty focusing is associated with both attention-deficit/hyperactivity disorder and major depressive disorder. A psychological test can place this

symptom within a broader context by comparing individual symptom patterns to a normative sample, thereby helping practitioners more accurately diagnose the relevant disorder. They can also provide insight into the etiology of specific disorders for a child, which can assist in both diagnosis and treatment planning.

Following diagnosis, psychological tests can also guide mental health treatment. First, psychological testing may implicate specific causal factors, which in turn help differentiate what evidence-based treatments may be most effective (Hunsley & Mash, 2020). For example, the etiological factors associated with depressive symptoms are diverse. Within this context, a psychological test may implicate relational stressors in the development of an individual's depressive symptoms; thus, the psychological test may help a clinician decide to use interpersonal psychotherapy. Similarly, psychological tests evaluate the contextual domains (e.g. social/interpersonal factors) that may mediate the efficacy of specific treatment approaches. For example, the Personality Assessment Inventory™–Adolescent contains questions assessing the adolescent's willingness to engage in treatment, the existence of social support, and positive impression management. Further, psychological testing can help practitioners focus on the most impairing symptoms as identified by the assessment (Hunsley & Mash, 2005). Overall, psychological tests provide insight into the lived experiences of clients and help shape the course of treatment using a well-established evidence-base.

Finally, psychological tests can assist in treatment monitoring and evaluation. Treatment monitoring includes tracking symptoms and functioning over time (Mash & Hunsley, 2005). Psychological tests can be invaluable in treatment monitoring as they provide objective, quantifiable data with which to observe changes in functioning. Though infrequently

implemented, the extant literature indicates that using tests as a form of treatment monitoring and feedback enhances treatment outcomes (Jenson-Doss et al., 2018). Further, psychological tests serve as an important form of treatment evaluation in determining the effectiveness, validity, cost effectiveness, and customer satisfaction of specific therapeutic methods (Mash & Hunsley, 2005).

Mental Health Challenges in the Child Welfare Population

Psychological tests and assessment are useful in clarifying rates of psychological disorders within specific populations. For example, it is well-documented that youth involved in the child welfare system are more likely to experience mental health symptoms and disorders than those in the general population (see Pecora et al., 2009; Vasileva & Peterman, 2018 for review). The extant literature attributes increased prevalence rates to higher-than-average exposure to complex trauma (e.g., exposure to multiple and/or varied traumatic experiences) and chronic trauma (e.g., repeated and prolonged exposure to maltreatment/violence; Kisiel et al., 2009), most notably child maltreatment as well as the often related experiences of parental dysfunction, intimate partner violence, and substance use disorders (Tanaka et al., 2011).

Additionally, this population often experiences ongoing complex trauma experiences because of contact with the child welfare system that also contribute to increased rates of mental health symptomology. It is impossible to consider the mental health of youth outside of the child welfare system's reinforcement of systemic oppression, as evidenced by the overrepresentation of families of color from low socioeconomic backgrounds in the system (Detlaff & Boyd, 2021). Indeed, early iterations of child welfare system functioned to police and pathologize families in order to increase the regulation of typically Black and Indigenous families (Mack, 2021). The

fruit of this history remains evident in child welfare policies today. Merkel-Holguin (2022) proposed that ongoing disproportionately is the result of child welfare policies designed around the most egregious cases of child abuse that actually function to criminalize poverty-related neglect, even while failing to address the systemic challenges facing minoritized groups.

Contact with these child welfare policies often in turn exposes children and families to traumatic experiences unique to the system. For example, these youth often experience significant ecological disruptions in their microsystem (Hong et al., 2011). Specifically, they are often separated from parents and other important support figures (e.g., family friends, religious leaders, teachers) when they are placed in out-of-home care (Stukes Chipungu & Bent-Goodley, 2004). The disruption of relationships in the microsystem is associated with increased likelihood for developing internalizing symptoms and externalizing behaviors (Berrick, 2006; Stukes Chipungu & Bent-Goodley, 2004), as well as substance use disorders (Keller et al., 2010). In addition, racism and classism have been documented as being embedded at every stage of the child welfare system in which families participate (Fluke et al., 2011). Specifically, families of color are more likely to be indicated for abuse and/or neglect (Rolock & Testa, 2005), are less likely to receive resources and support (Chibnall et al., 2003), and have longer lengths of stay in out-of-home care than white youth (Wulczyn et al., 2008). In this way, the child welfare system acts as an intensified microcosm of systemic injustice in the United States and magnifies the negative effect of these experiences on youth mental health.

In this context, the child welfare population is approximately four times more likely to meet criteria for a mental health disorder (Bronsard et al., 2016; Romano et al., 2018). The rates of specific symptoms and disorders also differ among children and youth in child welfare.

Among adolescents in the general population, anxiety disorders are the most commonly experienced form of psychopathology (Kessler et al., 2013); approximately one in eight adolescents meet clinical criteria for an anxiety disorder at any one point in time (Rockhill et al., 2010) while approximately 1 in 20 meets criteria for a depression or other mood disorder at any one time (Tang & Pinsky, 2015). In contrast, youth in contact with the system are three times more likely to experience clinically significant anxiety and four times more likely to experience depression (Bronsard et al., 2018). Further, Bronsard and colleagues (2016) meta-analytic review estimated that approximately 4% of youth involved in the child welfare system meet diagnostic criteria for posttraumatic stress disorder (PTSD); however, this number does not capture the many more youths in this population who experience clinically significant and impairing trauma symptoms who do not meeting diagnostic criteria for PTSD (Barboza & Dominquez, 2017).

Additionally, externalizing behaviors and associated disorders are the most prominent mental health concerns reported among youth in the child welfare system. Specifically, conduct disorder and oppositional defiant disorder have estimated prevalence rates of 3.5% and 2.8% respectively in the overall population (O'Connell et al., 2009), but children and youth in the child welfare system are 10 times more likely to carry this diagnosis (Bronsard et al., 2018; Moylan et al., 2010; Merikangas et al., 2011). These symptoms may be secondary to misdiagnosed or underdiagnosed developmental trauma. Further, where 7.2% of youth in the general population have a diagnosis of attention-deficit/hyperactivity disorder (ADHD) (Thomas et al., 2015), youth in the child welfare system are three times more likely to carry an ADHD diagnosis (Bronsard et al., 2018). Early externalizing behaviors can have longer-term consequences as well; the

presence of externalizing behaviors increase risk for developing substance use disorders (Brinkman et al., 2014).

In sum, because of prior and ongoing complex traumatic experiences, youth in the child welfare symptom are more likely to experience clinically significant internalizing symptoms and externalizing disorders as compared to youth in the general population. However, despite the well-developed body of literature underlining these increased prevalence rates, mental health symptomology among youth involved in the child welfare system often goes undetected. This may be the result of several factors. First, as previously noted, children and families in the child welfare system are more likely to experience systemic oppression as a result of racism and classism, which increase rates of complex and chronic mental health challenges. Additionally, youth in the child welfare system may be more likely to experience comorbid disorders, as youth in contact with the child welfare system have often experienced traumatic events that activate the neurobiological systems implicated in the development of psychopathology (Pecora et al., 2009). Because of this comorbidity, diagnosis and subsequent treatment are often more challenging in child welfare settings. For example, posttraumatic stress disorder and attention-deficit/hyperactivity disorder have several overlapping symptoms (e.g., difficulty concentrating, concerns with restlessness and executive functioning), but often have differing treatment courses (e.g., trauma-focused cognitive behavioral therapy versus parent management training). Further, youth in the child welfare system may experience greater discontinuity of care as compared to youth in the general population. Youth who move between placements may be less likely to attend to the same service providers and may therefore experience mental health symptoms that go undetected. As a result, child welfare stakeholders have advocated for the availability

psychological testing and assessment to provide for the standardized, rigorous identification of needs for this highly mobile population.

Multi-informant Psychological Testing

Effective interventions require the accurate assessment of the presence and severity of symptoms (Jenson-Doss & Hawley, 2010). The child welfare system typically uses both broadband and narrow assessment instruments to capture and quantify symptoms among children involved in the system. Child welfare case workers are encouraged to gather data from several reports (Capacity Building Center for States, 2017), as emerging evidence suggests that prioritizing assessment (i.e., integrating multiple sources of information) over individual psychological tests more effectively captures the mental health experiences of children and youth (De Los Reyes et al., 2019). Reporters may include parents, foster caregivers, teachers, and the adolescents themselves, as each may provide unique perspectives on the strengths and challenges associated with the adolescent and their family of origin.

However, despite the benefit of utilizing several perspectives, multiple informants do not inherently agree in their reporting. This has led to concerns by some about the practical validity of these assessments in general. However, others have argued that this is exactly why multi-informant assessments are necessary; they provide clinical utility to know how the child and close adults are perceiving symptom (De Los Reyes et al., 2015) and inform intervention work by suggesting reasons there are differences (e.g., low levels of trust in the relationship, differences across settings).

The extant literature has consistently observed low-to-moderate (i.e., r =.2-to-.3) agreement across informants (De Los Reyes, 2015). Patterns of multi-informant discrepancies in

the child welfare population were commensurate with the observed patterns in the broader population (i.e., r = .2-to-.3), though there is evidence to suggest that caregiver-child reports are more likely to converge among youth with maltreatment histories (Romano et al., 2018). Makol et al. (2020) summarized three patterns agreement/disagreement across studies examining multiple reporters: (1) reporters may agree on low levels of symptoms, (2) reporters may agree on high levels of symptoms, (3) or one informant endorses symptoms not reported by another. These patterns were consistent regardless of reporter dyad (e.g., caregiver-teacher, caregiver-child, child-teacher) and presenting problem (e.g., externalizing vs. internalizing disorder). However, there are trends in level of disagreement specific to who is reporting on the child's mental health. Indeed, informant pairing is one of the most important factors in determining level of disagreement (der Ende et al., 2012).

Multi-informant reporters are more likely to correspond when observing symptoms in the same context (De Los Reyes, 2015; Achenbach et al., 2019). Consistently, dyadic reporters from the same context (e.g., mother and father) yield more consistent symptom ratings than those from different settings (e.g., mother and teacher; De Los Reyes, 2015; Makol et al., 2020). Specifically, caregivers tend to report more internalizing symptoms and externalizing behaviors than teachers (Rescorla et al., 2011) and teachers tend report fewer internalizing and externalizing problems than caregivers and youth (Youngstrom et al., 2000), though teacher reports also capture important psychosocial functioning information.

However, though youth may be reporting on that same setting as other informants (e.g., school functioning), there are significantly more discrepancies across youth-adult reports than adult-adult reports. In the general population, lowest levels of informant agreement are typically

observed when comparing the youth's report of their own mental health functioning with that of an adult (r = .22; Martel et al., 2017). Broadly, youth tend to report higher levels of psychopathology than their caregivers (Burk & Laursen, 2010) and their teachers (Youngstrom et al., 2000).

The picture is more complicated in child welfare settings given that multiple caregivers are often involved. For example, teachers and kinship caregivers (e.g., relative caregivers) tend to have higher levels of convergence as compared to teachers and non-relative foster parents, as non-relative foster caregivers are more likely to report higher levels of problem behavior at home (Shore et al., 2002). In addition, foster caregivers still frequently under- and overestimate the difficulties that youth experience (McWey et al., 2014) and are more likely to correspond with teacher reports than with youth reports (Makol et al., 2020). Despite these patterns, Parker et al. (2019) found that youth converge on reports of internalizing symptoms and externalizing behaviors more closely with foster caregivers than with biological parents. Differences in reporting are also found between maltreating and non-maltreating caregivers (Romano et al., 2018).

There are several factors that may contribute to this discrepancy. For example, biological caregivers are often fearful about the implications of the screening or fear loss of custody resulting from reporting on youth's mental health needs (Darlington et al., 2005). Additionally, families may have an unfavorable view of services, including those related to mental health, because of their prior experiences with child welfare services (Parker et al., 2018) and the child welfare system's history of upholding white supremacy (Mack, 2021). This mistrust may be further compounded by stigma related to mental health in Black and Latina/e/o communities

(Alvidrez et al., 2008). Thus, there are several complex incentives that may contribute to the pattern of underreporting observed in biological parents involved in the child welfare system.

In addition to multi-informant pairing, discrepancies are also related to symptom type. Specifically, though caregivers consistently rate symptoms as more problematic than teachers (Rescorla et al., 2014), multi-informant reports are more likely to agree when reporting on externalizing behaviors (McWey et al., 2014). The extant literature has regularly found higher levels of convergence across reports of externalizing behavior. This pattern is true when examining child-teacher and teacher-caregiver pairings, including those containing non-relative foster parents (Romano et al., 2018), though teacher-caregiver reports have higher levels of agreement when compared to child-caregiver reports (Makol et al., 2020). This is likely because externalizing behaviors are easier to observe. Reporters may also be sensitized to noticing and reporting the presence of externalizing behaviors due to biases associated with knowing a child's maltreatment and/or child welfare status. In contrast, reporters are more likely to disagree when describing youth with comorbid internalizing symptoms and externalizing disorders (Hoffman & Chu, 2015).

There are also several child demographic factors that influence discrete patterns of multi-informant reporting discrepancies across informant dyads. First, gender may influence level of agreement; however, the way gender may be associated with discrepancies and why remains unclarified. When examining caregiver reports only, caregivers are more likely to report internalizing symptoms in girls and externalizing behaviors in boys (Tarren-Sweeney et al., 2004), which may be the result of socialization and gendered expectations (Rosenfield, 2000). When examining discrepancies across reporters, McWey et al. (2018) found that there was

typically less agreement between caregivers and youth when the adolescent is male. Regarding teacher reports, prior literature indicates that teachers are more likely to perceive oppositional behavior, inattention, and hyperactivity in males (Jackson & King, 2004). No literature has yet examined reporting discrepancies between teacher-caregiver and child-teacher dyads regarding symptom cluster. However, De Los Reyes & Kazdin (2005) found that while correspondence among informants may be influenced by gender in specific settings (e.g., clinic-, community-, and school-based settings), aggregated data across populations does not indicate significant gender effects on level of reporter discrepancy. Thus, it is important to examine whether the child welfare setting contributes to gender-based patterns of reporter discrepancy to better inform the interpretation of these assessments in casework practice.

Additionally, the extant literature indicates that race/ethnicity is related to differences in reporting between parents and teachers (Rescorla et al., 2013; Lawson et al., 2017), which may be the result of teacher's racism (Kang & Harvey, 2020). Relatedly, child-teacher dyads tend to be more discrepant when the adolescent is male and Black as opposed to white (Youngstrom et al., 2000). In terms of parent-child dyads, Lau and colleagues (2004) found that parents from systemically oppressed backgrounds tend to report fewer internalizing symptoms and externalizing behaviors as compared to youth, whereas Caucasian dyads are more likely to have parents reporting higher levels of symptoms. It is likely that parents are also experiencing systemic oppression, and this may affect their reporting of mental health challenges (Pactor & Garcia Coll, 2009).

In terms of age, while reporter discrepancies occur across the lifespan (De Los Reyes et al., 2019), informants are more likely to correspond when reporting on younger children as

compared to adolescents (De Los Reyes, 2015; Smith, 2007), which prior research has attributed the relatively limited number of contexts in which younger children can be observed (De Los Reyes, 2015). Additionally, adolescence is defined by increasing autonomy from and conflict with the family unit (Branje, 2018) and increasing reliance on peer relationships (Blakemore & Mills, 2014); as such, adolescents may be less likely to confide in caregivers or teachers about their mental health challenges, which in turn mitigates the ability of reporters to accurately describe a teenager's mental health symptoms. Thus, it is increasingly important to capture youth's perspective on their mental health functioning as they age in order to increase diagnostic clarity (Smith, 2007).

Further, child factors related to child welfare involvement also influence discrete patterns of multi-informant reporting. Specifically, caregivers are more likely to report internalizing symptoms among children and adolescents who have experienced complex trauma including physical abuse, neglect, and sexual abuse as compared to youth who have not but are more likely to report externalizing behaviors when a youth has experienced only physical abuse (McWey et al., 2018). While the extant literature has not examined how teacher reports are influenced by child maltreatment history, Martin and colleagues (2010) found that teachers commonly believe that maltreatment manifests in attention challenges and disruptive behaviors. In sum, while informant pairing and problem type are the most salient predictors of multi-informant discrepancies, child demographic factors like age, gender, race/ethnicity, and maltreatment history contribute to patterns of reporting.

Clinical Importance of Multi-Informant Discrepancies

While discrete patterns of divergence emerge across problem and informant types, it does not necessarily follow that these differences are unimportant. Indeed, the discrepancy across reporters often yields meaningful diagnostic data (Strijker et al., 2011). Multiple reporter discrepancies may capture symptom differences across settings (De Los Reyes et al., 2015). Specifically, reporters may in fact be discerning different symptomology depending on their environment of observation as youth may exhibit symptoms in systematically variable ways across settings (Achenbach et al., 2019). These variations may provide essential diagnostic information. For example, a child may be more likely to exhibit the hyperactive symptoms of attention-deficit/hyperactivity disorder (ADHD) in school settings where the child's attention and executive functioning are more greatly taxed as compared to the home environment. In this situation, teacher reports of symptoms are essential in providing diagnostic clarity about impairment across domains and insight into what services are needed to support the child's academic functioning. The Diagnostic and Statistical Manual reflects this complexity: symptoms of ADHD must be apparent across two or more settings (American Psychological Association, 2013). Thus, discrepancies across reports may have profound diagnostic implications.

Next, discrepancies may capture relationship differences across reporters (van Dulmen & Egeland, 2011; Connelly & Ones, 2010). Informant ratings may illustrate differences in how the adolescent behaves with the informant (Achenbach et al., 2019). For example, adolescent-caregiver closeness mediates observed discrepancies across multiple reporters (McWey et al., 2018). Thus, disagreement may indicate lower levels of relational closeness. In this way, discrepancies due to relationship differences may be particularly important to conceptualize in

child welfare settings as they may help inform case practice. Specifically, differences in relational intimacy, warmth, trust, and communication may help determine best fit in terms of placement or help shape what services are needed to build to reunification. Thus, multiple reporters are valuable for accurately assessing and diagnosing mental health problems among adolescents and in understanding related treatment factors.

There is also some evidence to suggest that multi-informant reporting discrepancies are also related to mental health challenges above and beyond the negative outcomes associated with psychiatric disorders that onset during adolescence (Keller et al., 2010). Among caregiver-child dyads, De Los Reyes and Kazdin (2005) and De Los Reyes (2013) proposed a model wherein different adolescent and parent attributions regarding the cause of mental health symptoms contribute to differences in clinical practice. As there are no clinical guidelines for combining multi-informant reports, clinicians may preference one report over another based on their own judgment. However, focusing on one report lowers the sensitivity of diagnostic assessment and screening protocols (Thakur & Cohen, 2019) and contributes to incorrect diagnoses (Achenbach et al., 2019). This may in turn contribute to the associated negative outcomes (e.g., child behavioral and emotional problems, poor behavioral, work, and criminal outcomes; De Los Reyes, 2011) that are associated with caregiver-child reporting discrepancies.

Discrepancies are also associated long-term challenges regarding symptom management and service utilization. For example, caregiver-child discrepancy is associated with poorer treatment outcomes, particularly when the caregiver reports higher symptomology, regardless of symptom type (Goolsby et al., 2018). This may be because children are less motivated to engage in treatment when they do not believe there is a problem, and suggests the need for ongoing

psychoeducation, motivational interviewing, and interventions designed to facilitate trust between the caregiver and teen. Relatedly, among child-teacher dyads, there is evidence to suggest that child-teacher reporting discrepancies are associated with fewer treatment improvements among youth with ADHD diagnoses (Hennig et al., 2017).

There is also evidence to suggest that adolescent-caregiver disagreement in reporting is associated not only with continued symptoms, but with increases in symptoms over time (Ferdinand et al., 2004), including among child welfare populations (McWey et al., 2014). For example, Augenstein et al. (2021) found an increased likelihood of future suicidal ideation among adolescents when youth reported elevated depressive symptoms and caregivers reported subthreshold depressive symptoms. This may be the result of poor communication and family functioning between caregivers and adolescents, poor parental insight into existing problems, challenges secondary to mental health stigma, or poorer treatment engagement. It may also be the result of assessment measures that are not sensitive enough to assess for culturally mediated depression symptoms, particularly as many assessment measures were norm on white middle- to upper-class youth.

Despite evidence suggesting that child-caregiver and child-teacher reporting discrepancies are associated with poorer long-term mental health outcomes and/or less service utilization, there is a relative dearth of literature examining the relationship between caregiver-teacher dyadic reporting discrepancies. One notable exception includes a study by Ferdinand and colleagues (2006), who examined the relationship between caregiver-teacher reporting discrepancies and long-term child mental health outcomes. They found that while caregiver and teacher ratings predicted future mental health functioning, the discrepancy between reports did

not. However, Ferdinand and colleagues (2006) did not examine how these discrepancies changed between first assessment and follow up. This is a significant gap as there is evidence to suggest that discrepancies are not static; rather, they change in response to relational, contextual, and treatment factors.

For example, Becker-Haimes and colleagues (2018) found that informant discrepancies decreased over the course of treatment when parents reported more internalizing symptoms at time one. Further, an der Ende et al. (2012) examined how self-parent, caregiver-teacher, and teacher-self reporting discrepancies vary across the lifespan. They found that discrepancies increase as the child ages across dyads. Additionally, these changes differ with reference to symptom cluster. Specifically, discrepancies between self-parent and caregiver-teacher dyads increases with age when rating internalizing symptoms and decreases with age when rating externalizing behaviors. This is consistent with the extant literature suggesting that ratings of externalizing symptoms are associated with higher rates of correspondence. Thus, it is important to examine how caregiver-teacher discrepancies change over time and relate to mental health functioning. Correspondence between these reporters may be particularly salient for youth previously or currently involved in the child welfare system as teachers often function as key members of the social support network. Specifically, discrepancies between caregivers and teachers may represent discord and/or low levels of communication, which may negatively affect the child's experience of social support and the child welfare case plan. Thus, more work is needed to examine the relationship between caregiver-teacher reporting discrepancies and their relation to long-term mental health outcomes.

Despite the contributions of several important studies, the extant literature remains piecemeal. For example, studies have often focused on specific mental health disorders, such as depression, rather than broad symptom clusters (e.g., internalizing symptoms and externalizing behaviors). The extant literature also has prioritized research examining caregiver-child and teacher-child reporting discrepancies. Indeed, no study to date has examined caregiver-teacher reporting discrepancies, the trajectory of change associated with these discrepancies over time, and related mental health outcomes in later adolescence among the United States child welfare population. This is particularly concerning as the importance of mental health assessments and their interpretations by child welfare professionals cannot be overstated.

Case decisions are heavily influenced by caseworker perception (Berger et al., 2010), particularly as high caseloads may limit the ability of child welfare professionals to collect adequate information or to effectively analyze discrepancies among that information (Capacity Building Center for States, 2017). Thus, multi-informant discrepancies have profound implications for case planning and decision making in child welfare contexts. For example, adolescents may be more likely to be placed into a more restrictive setting (e.g., therapeutic school) if caseworkers tend to prefer foster caregiver reports, which, as previously stated, tend to overemphasize adolescent psychopathology. Placement in a more restrictive setting than their mental health needs necessitate is associated with slower improvement gains (Chor et al., 2014). Placement in a more restrictive setting is also considered a “non-progress” move in terms of treatment goals (Font et al., 2018) and may extend time that an adolescent is in foster care, which is associated with increased mental health symptoms (Vasileva & Peterman, 2018), as well as lower educational achievement (O’Higgins et al., 2017), poorer physical and behavioral well-

being (Villodas et al., 2016), and increased likelihood of substance use disorders and criminal involvement (Gypen et al., 2017). Alternately, if caseworkers make decisions based on underreported mental health symptoms, youth may be unable to access necessary services. For example, if someone rates symptoms differently than another, a residential placement team may be less likely to recommend this higher level of care when it is actually indicated. Thus, while multi-informant discrepancies may have profound effects on the experiences of youth in the child welfare, very little research has examined how these discrepancies contribute to long-term outcomes among this highly vulnerable population.

The Current Study

Thus, the current study seeks to address this gap. First, it examines reporting discrepancies among caregivers, teachers, and 12-year-old adolescent dyads with prior and/or ongoing child welfare involvement. Ratings of both internalizing symptoms and externalizing behaviors were collected using the Child Behavior Checklist (CBCL), Teacher Report Form (TRF), and the Youth Self-Report (YSR), which have been previously validated as comparable measures (van Dulmen & Egeland, 2011). Teacher-caregiver discrepancies were also assessed at ages 10, 14, and 16. Next, trajectories based on teacher-caregiver discrepancies were assessed using growth mixture modeling to identify trajectories of teacher-caregiver discrepancies when rating internalizing symptoms, externalizing behaviors, and total mental health symptoms respectively. Finally, the current study examines how the multi-informant discrepancy trajectories predict adolescent mental health self-reporting at age 18. Taken together, the current study addresses three primary research questions:

- **Research question 1:** What discrepancies exist in and between caregiver-child, child-teacher, and teacher-caregiver reports of adolescent mental health symptomology at age 12? How do these discrepancies relate to the demographic variables of gender and ethnicity?
    - **Hypothesis 1.1:** Caregiver-child reports will have smaller magnitudes of discrepancy as compared to child-teacher discrepancies.
    - **Hypothesis 1.2:** Caregiver-teacher reports will have smaller magnitudes of discrepancy as compared to child-teacher discrepancies.
    - **Hypothesis 1.3:** Caregiver-child dyads will have smaller magnitudes of discrepancy when reporting on internalizing symptoms at age 12 when the adolescent is female.
    - **Hypothesis 1.4:** Caregiver-child dyads will have smaller magnitudes of discrepancy when reporting on externalizing behaviors at age 12 when the adolescent is male.
    - **Hypothesis 1.5:** Caregiver-teacher dyads will have smaller magnitudes of discrepancy when reporting on internalizing symptoms at age 12 when the adolescent is female.
    - **Hypothesis 1.6:** Caregiver-teacher dyads will have smaller magnitudes of discrepancy when reporting on externalizing behaviors at age 12 when the adolescent is male.

- **Hypothesis 1.7:** Child-teacher dyads will have smaller magnitudes of discrepancy when reporting on internalizing symptoms at age 12 when the adolescent is female.
- **Hypothesis 1.8:** Child-teacher dyads will have smaller magnitudes of discrepancy when reporting on externalizing behaviors at age 12 when the adolescent is male.
- **Hypothesis 1.9:** Caregiver-child dyads will have smaller magnitudes of discrepancy when reporting on total mental health symptoms at age 12 when the adolescent is Black.
- **Hypothesis 1.10:** Child-teacher dyads will have higher magnitudes of discrepancy when reporting on total mental health symptoms at age 12 when the adolescent is Black.
- **Hypothesis 1.11:** Teacher-caregiver higher will have smaller magnitudes of discrepancy when reporting on total mental health symptoms at age 12 when the adolescent is Black.

- **Research question 2:** How do reporter discrepancies between caregiver-teacher dyads change over adolescence?
  - **Hypothesis 2.1:** A four-class solution will be the best trajectory fit for caregiver-teacher ratings of internalizing symptoms and will include 1) a class with consistently high levels of discrepancy, 2) a class with consistently low levels of discrepancy, 3) a class with increasing levels of discrepancy, and 4) a class with decreasing levels of discrepancy.

- **Hypothesis 2.2:** A four-class solution will be the best trajectory fit for caregiver-teacher ratings of externalizing behavior and will include 1) a class with consistently high levels of discrepancy, 2) a class with consistently low levels of discrepancy, 3) a class with increasing levels of discrepancy, and 4) a class with decreasing levels of discrepancy.
- **Hypothesis 2.3:** A four-class solution will be the best trajectory fit for caregiver-teacher ratings of total mental health symptomology and will include 1) a class with consistently high levels of discrepancy, 2) a class with consistently low levels of discrepancy, 3) a class with increasing levels of discrepancy, and 4) a class with decreasing levels of discrepancy.

- **Research question 3:** How do caregiver-teacher discrepancy trajectories predict child mental health ratings at age 18?

  Internalizing Symptoms

  - **Hypothesis 3.1**: Internalizing symptom trajectories associated with lower levels of disagreement between caregivers and teachers in either direction will predict lower child-rated internalizing symptoms at age 18.
  - **Hypothesis 3.2**: Internalizing symptom trajectories associated with higher levels of disagreement between caregivers and teachers in either direction will predict higher child-rated internalizing symptoms at age 18.
  - **Hypothesis 3.3**: Internalizing symptom trajectories associated with higher levels of disagreement between caregivers and teachers in either direction will predict higher child-rated internalizing symptoms at age 18.

  Externalizing Behaviors

- **Hypothesis 3.4**: Externalizing symptom trajectories associated with lower levels of disagreement between caregivers and teachers in either direction will predict lower child-rated Externalizing symptoms at age 18.
- **Hypothesis 3.5:** Externalizing symptom trajectories associated with higher levels of disagreement between caregivers and teachers in either direction will predict higher child-rated externalizing symptoms at age 18
- **Hypothesis 3.6:** Externalizing symptom trajectories that change over time regarding level of disagreement between caregivers and teachers will predict significantly higher child-rated externalizing symptoms at age 18.

Total Symptoms

- **Hypothesis 3.7**: Total symptom trajectories associated with lower levels of disagreement between caregivers and teachers in either direction will predict lower child-rated total symptoms at age 18.
- **Hypothesis 3.8**: Total symptom trajectories associated with higher levels of disagreement between caregivers and teachers in either direction will predict higher child-rated total symptoms at age 18.
- **Hypothesis 3.9:** Total symptom trajectories that change over time regarding level of disagreement between caregivers and teachers will predict higher child-rated total symptoms at age 18.
- **Hypothesis 3.10:** Total symptom trajectories where teachers always rate symptoms higher than caregivers will be associated with higher levels of child-reported total symptoms at age 18.

CHAPTER THREE

METHODS

Participants

Data for the current study were drawn from the Longitudinal Studies in Child Abuse and Neglect (LONGSCAN) Consortium of Studies of Child Abuse and Neglect (Runyan & Kotch, 2014). Founded in 1990, the LONGSCAN Consortium was composed of five data collection sites located in the East (n = 282), Midwest (n = 245), South (n = 243), Southwest (n = 330), and Northwest (n = 254). The consortium assessed over 1,300 children, their caregivers, and teachers in two-year increments form ages four to 20 years old. Data related to maltreatment were also collected from Child Protective Services at two-year intervals.

The present study includes a total of 303 participants. The sample was restricted to include only youth with prior and/or ongoing child welfare involvement. Thus, 10.56% (n = 32) were recruited from the Midwest, 36.30% (n = 110) from the Northwest, and 53.14% (n = 161) from the Southwest. The main effects of site on the relevant controlling and dependent variable were compared. There were no significant differences observed across sites regarding gender ($X^2(2) = 1.39$, $p = 0.05$). A significant number of participants did not have race/ethnicity data reported; thus, sites were not compared regarding this demographic variable. Given limited collinearity between sites and variables of interest, the sites are equivalent and thus able to be compared.

The sample includes slightly more girls (52.15%; n = 158) than boys (48.85%; n = 145). Approximately 15% of participants were missing data related to ethnicity. In terms of ethnicity,

18.15% (*n* = 55) were Black, 16.50% (*n* = 50) were White, 12.54% (*n* = 38) Mixed/Multiracial, 7.26% (*n* = 22) Indigenous peoples, 2.64% (*n* = 8) Latina/e/o, and 1.32% (*n* = 4) Asian; 41.58% (*n* = 126) were missing data.

Procedure

The Midwest (sampled in Chicago) site recruited participants from those who were referred to child protective services; the Northwest (sampled in Seattle) included children who were reported to Child Protective Services and believed to be at moderate risk for subsequent maltreatment; the Southwestern (sampled in San Diego) site included participants placed into out-of-home care as a result of maltreatment during the first 42 months of life; the Southern (sampled in North Carolina) sites included participants "at risk" of maltreatment as indicated by a screening measure; the Eastern (sampled in Baltimore) site included children from low-income families recruited from primary health care clinics. The present study included participants from the Midwest, Northwest, and Southwest sites. Thus, while participating families differ in their history of substantiated maltreatment, all have been in contact with the child welfare system.

After recruitment, participating children, caregivers, and teachers completed a series of assessments at 4, 6, 8, 10, 12, 14, 16, 18, and 20 years of age for the child. Assessments included in-person interviews among children and caregivers, as well as a number of mailed surveys and questionnaires completed by children, caregivers, and teachers. Brief telephone encounters were conducted between assessment years to enhance retention and document any significant life events.

Measures

Internalizing Symptoms, Externalizing Behaviors, and Total Symptoms. The Child Behavior Checklist (CBCL/4-18; Achenbach, 1991) is a widely used broadband instrument used

to assess childhood psychopathology. The CBCL was normed through a multistage probability sample from 100 sites across the contiguous United States (Achenbach & Rescorla, 2001), which makes it particularly appropriate for the scale of LONGSCAN data collection. A parent or guardian rates 118 items on a 3-point Likert scale where higher scores represent the presence of more child behavior problems (0 = not true, 1 = somewhat or sometimes true, 2 = very true or often true). Responses to these questions inform ratings across eight problem-specific scales (i.e., social withdrawal, somatic complaints, anxiety/depression, social problems, thought problems, attention problems, delinquent behaviors, and aggressive behaviors), which are each included in the total problems summary scale. Additionally, an internalizing symptoms summary scale is composed of the anxious/depressed, withdrawal, and somatic complaints problem-specific scales. An externalizing symptoms summary scale is composed of the delinquent behaviors and aggressive behaviors syndrome scales. The current study utilizes CBCL data from the parent or guardian respondents rating the child's behavior at ages 10, 12, 14, and 16.

Child ratings of mental health symptomology were obtained using the Youth Self-Report (YSR) and teacher ratings of child mental health symptomology will be obtained using the Teacher Report Form (TRF; Achenbach & Rescorla, 2001). The YSR and TRF are parallel measures to the CBCL containing 112 and 113 items respectively. Questions are answered using a 3-point Likert scale (0 = not true, 1 = somewhat or sometimes true, 2 = very true or often true). Like the CBCL, the higher the raw score the higher the number of child behavior problems. Responses are used to form the same eight problem-specific scales and three summary scales as those derived from the CBCL. The current study will be utilizing teacher ratings of child mental health at ages 10, 12, 14, and 16. Child ratings of their mental health functioning will be assessed using the YSR at age 12 and as an outcome variable at age 18.

The CBCL, TRF, and YSR are each components of the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach, 2009) and have been used in over 9,000 empirical studies (Rosanbalm et al., 2016). The CBCL, TRF, and YSR have demonstrated internal consistency, test-retest reliability, construct validity, and criterion related validity in relation to the clinical scales. This is true across the general and clinical populations, as well as among children and youth who have experienced trauma (Rosanbalm et al., 2016). Thus, ASEBA assessment measures are commonly used among the child welfare population (e.g., Woods et al., 2013). Although the extant literature indicates that there are limitations to comparing the measures consistent with the limitations in multi-informant assessments more broadly (Gomez et al., 2014), the CBCL, TRF, and YSR have been used in numerous studies comparing child, teacher, and caregiver ratings of mental health and have demonstrated convergent validity across settings and populations (Grigorenko et al., 2010; Achenbach et al., 2008) as they have analogous items, subscales, and factor structures (Achenbach & Rescorla, 2001).

Demographic Data. Information related to the child's gender and ethnicity was captured via the Multi-group Ethnic Identity Measure, which was completed by youth respondents at age 12 during the child interview. This measure has demonstrated reliability among a high school sample (LONGSCAN Manual).

Data Analyses

Preliminary Analyses. To compare across measures, raw scores were converted into z-scores. In addition, when multiple teachers completed TRFs at one timepoint, the mean of the two scores included in order to incorporate both perspectives into comparisons without allowing

one individual to have two to three datapoints included in the larger analysis. This process was repeated for caregiver and teacher data collected at ages 10, 14, and 16.

Initial analyses were conducted to ensure that data collected from each of the three LONGSCAN sites were comparable. Correlations were conducted to determine the extent to which CBCLs and TRFs were similar across sites. Next, ANOVAs were conducted to compare differences between guardian's CBCL responses and teacher's TRF responses ages 10, 12, 14, and 16.

Aim One: Discrepancies Between Caregivers, Teachers, and Children at Age 12. The extant literature suggests that the direction of the discrepancy (e.g., who reports more symptoms) and the amplitude of the discrepancy (e.g., absolute differences of discrepancies between caregiver and child) are both important aspects of overall discrepancies. Thus, the current study will examine not only the direction the discrepancy score falls into, but also the distance from zero. To do so, separate numbers were derived to represent to direction and amplitude of discrepancy for the three subscales on the CBCL, TRF, and YSR (e.g., internalizing symptoms, externalizing behaviors, total symptomology). Specifically, a difference score was calculated to capture the magnitude of the discrepancy and then a categorical number was assigned to represent the direction (e.g., 1 = teacher reported higher, 2 = caregiver reported higher). Both numbers were then incorporated into a model to capture the overall discrepancy (e.g., amplitude and direction) at age twelve. *T*-tests were then conducted to determine how discrepancies differed by the child factors of gender and ethnicity.

Aim Two: Examining Caregiver-Teacher Discrepancy Trajectories. Growth mixture modeling was then used to identify latent trajectories related to level of caregiver-teacher dyadic discrepancies at ages 10, 12, 14, and 16. Because the extant literature has not examined distinct

longitudinal trajectories among this population, growth mixture modeling enabled the researcher to assess unobserved groups without *a priori* hypotheses programmed into the model. Further, growth mixture modeling elucidated how trajectories differ with respect the mean amount of change, interindividual differences, and pattern of change over time (Ram & Grimm, 2009).

Plotted trajectories were then examined to determine adequate heterogeneity of discrepancy changes over time (i.e., spaghetti plot). Given significant observed heterogeneity, a one-class growth curve model was fitted to the data to serve as a baseline for the unobserved groups before a series of models were examined to determine the best single-group representation of change (Ram & Grimm, 2007).

After this best fitting unconditional model was fit, the number of subgroups examined was determined using practical and theoretical insights. Specifically, as previously mentioned, the extant literature suggests that there are three distinct patterns of caregiver-child reporting discrepancies: (1) reporters may agree on low levels of symptoms, (2) reporters may agree on high levels of symptoms, (3) or one informant endorses symptoms not reported by another (Makol et al., 2020). Additionally, McWey et al. (2015) found that the direction of disagreement (i.e., caregiver reporting more symptoms than child, vice versa) represented distinct groups. Thus, the current study predicted four identifiable subgroups. Thus, growth mixture models were fit increasing from the base model fit to the determined maximum ($k = 5$).

Model fit was evaluated using multiple fit information criteria. Specifically, entropy, which represents overall confidence in class assignment, was used as a guiding fit index given that entropy values greater than >0.80 are optimal for interpretation. Addition fit criterion, including Bayesian Information Criterion (BIC), sample-size-adjusted BIC (ssBIC), Akaike's Information Criterion (AIC), and Consistent AIC (CAIC) were then examined to determine the

parsimony of each model. Priority will be paid to models with smaller values associated with each information criteria. Additionally, Classification likelihood criterion (CLC), Integrated Completed Likelihood Criterion with BIC approximation (ICL-BIC), Normalized Entropy Criterion (NEC), and Entropy (E). Smaller values for CLC, ICL-BIC, and NEC were preferred (van de Schoot et al., 2017). Of note, fit indices do not provide absolute clarity related to the number of "true" groups. Rather, they were used in conjunction with clinical judgment and guiding theory to determine the most appropriate relative fit among the models (Frankfurt et al., 2016).

Aim Three: Predicting Child Mental Health Ratings at Age 18. Class membership was then incorporated multiple regression analyses to predict child-rated mental health symptoms at age 18. To control for baseline child-rated mental health symptoms at age 12, covariates were entered into separate blocks to determine their independent effect on child-rated mental health outcomes at age 18. Of note, a significant number of participants were missing data regarding gender and race/ethnicity; as such, these variables were not included as covariates in multiple regression analyses.

CHAPTER FOUR

RESULTS

Preliminary Analysis

Several analyses were conducted to determine differences between rating measures across LONGSCAN sample sites. First, correlations between the CBCL and TRF were conducted for internalizing symptom (Table 1), externalizing behaviors (Table 2), and total symptom (Table 3) scores.

Table 1. *CBCL and TRF Internalizing Symptom Correlations Among Sites*

Ages	Full Sample	MW	NW	SW
10	.26**	0.30**	--	0.26**
12	.21**	0.28	0.22*	0.20*
14	.12*	0.01	0.15	0.12
16	0.29**	--	0.29**	0.34**

* < 0.05, ** < 0.01
-- Unable to be computed

Table 2. *CBCL and TRF Externalizing Behavior Correlations Among Sites*

Ages	Full Sample	MW	NW	SW
10	.43**	0.69**	--	0.43**
12	.39**	.47*	.25**	.49**
14	0.28**	0.28	0.29**	0.37**
16	0.23**	--	0.26**	0.21*

* < 0.05, ** < 0.01
-- Unable to be computed

Table 3. *CBCL and TRF Total Symptom Correlations Among Sites*

Ages	Full Sample	MW	NW	SW
10	.33**	.43	--	0.33**
12	.37*	.45**	.24**	.45**
14	.24**	.21	.25**	.28**
16	.28**	--	0.26**	0.30**

* < 0.05, ** < 0.01
-- Unable to be computed

In addition, ANOVA analyses were conducted to determine if CBCL and TRF internalizing, externalizing, or total symptom ratings differed across sites. At age 10, internalizing CBCL scores ($F(2, 292) = 2.35$, $p = 0.09$) or TRF scores ($F(1,124) = 0.151$ $p = 0.48$) did not significantly differ by site. While externalizing CBCL scores ($F(2, 292) = 0.2.67$, $p = 0.07$) did not differ, TRF scores ($F(1,124) = 5.476$, $p = 0.02$) did significantly differ by site. Post-hoc Tukey analyses indicated that children from the MW site were more likely to have higher externalizing behavior TRF scores as compared to the SW site. Finally, total symptom CBCL scores ($F(2, 292) = 2.63$, $p = 0.07$) or TRF scores ($F(1,124) = 2.51$, $p = 0.12$) also did not differ by site at age 10.

At age 12, neither internalizing CBCL scores ($F(2, 313) = 0.32$, $p = 0.73$) or TRF scores ($F(2, 270) = 0.16$, $p = 0.85$) significantly differed by site. Similarly, there was not a significant difference regarding externalizing CBCL scores ($F(2, 313) = 1.33$, $p = 0.27$) or TRF scores ($F(2, 270) = 0.46$, $p = 0.63$) across sites. Finally, total symptom CBCL scores ($F(2, 313) = 1.17$, $p = 0.31$) or TRF scores ($F(2, 270) = 0.05$, $p = 0.95$) also did not differ by site.

At age 14, internalizing CBCL scores ($F(2, 331) = 1.65$, $p = 0.19$) and TRF scores ($F(2, 283) = 1.69$, $p = 0.19$) did not significantly differ by site. While CBCL externalizing ratings did not differ across sites ($F(2, 331) = 2.57$, $p = 0.08$), TRF externalizing ratings did differ across sites ($F(2, 283) = 4.57$, $p = 0.01$). Post-hoc Tukey HSD analysis indicated that children in from

the MW sample had significant higher TRF scores as compared to those from the NW ($p = 0.01$) and SW ($p = 0.01$) samples. Finally, total symptom CBCL scores ($F(2, 331) = 2.84$, $p = 0.06$) or TRF scores ($F(2, 283) = 1.61$, $p = 0.20$) also did not differ by site at age 14.

At age 16, internalizing CBCL scores ($F(2, 303) = 2.35$ $p = 0.09$) and TRF scores ($F(1, 203) = 1.74$, $p = 0.19$) did not significantly differ by site. TRF externalizing behavior scores ($F(2, 203) = 0.001$, $p = 0.98$) and total symptom scores ($F(2, 203) = 0.26$, p = 0.61). However, there was a significant difference regarding externalizing CBCL scores ($F(2, 303) = 3.21$, $p = 0.04$) and total symptom CBCL scores did differ ($F(2, 303) = 3.24$, $p = 0.04$). Post-hoc Tukey HSD analyses indicated SW site had higher CBCL externalizing behavior scores ($p = 0.04$) and total symptom scores ($p = 0.03$) as compared to youth from the MW sample. Taken together, while minor differences are present across sample sites, the three sites are comparable on key variables and thus able to be utilized within a single sample.

Aim One

Interrater Correspondence. Correlations of the magnitude of rating discrepancy between caregiver-child, teacher-child, and teacher-caregiver dyads were calculated to examine significant differences between dyad disagreement (Table 4).

Table 4. *Correlations Among Dyads by Symptom Cluster*

	Symptom Cluster		
	Internalizing symptom *r*	Externalizing behaviors *r*	Total symptoms *r*
Informant dyad			
Caregiver-child	.32	.43	.38
Teacher-child	.25	.37	.25
Caregiver-teacher	.24	.33	.33

Magnitudes of Discrepancy Across Dyads at age 12. Paired sample *t*-tests were conducted to examine differences in magnitudes of discrepancy between dyads (Table 5). In contrast to hypothesis 1.1, there were no significant differences between magnitudes of discrepancy when rating internalizing symptoms ($t(113) = 0.09$, $p = .93$), externalizing behaviors ($t(113) = 0.36$, $p = .72$), or total symptoms ($t(113) = 0.82$, $p = .41$) between caregiver-child and child-teacher dyads. Similarly, in contrast to hypothesis 1.2, there were no significant differences between magnitudes of discrepancy when rating internalizing symptoms ($t(113) = 0.04$, $p = .97$), externalizing behaviors ($t(113) = 0.71$, $p = .48$), or total symptoms ($t(113) = 0.11$, $p = .92$) between caregiver-teacher and child-teacher dyads.

Table 5. *Absolute Magnitudes of Discrepancy (Z-scores) by Symptom Cluster*

	Internalizing Symptoms		Externalizing Behaviors		Total Symptoms	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Caregiver-child	0.87	0.75	0.80	0.61	0.85	0.68
Teacher-child	0.86	0.84	0.83	0.72	0.92	0.72
Caregiver-teacher	0.85	0.78	0.88	0.74	0.93	0.71

Effect of Gender on Magnitudes of Discrepancy Across Dyads at age 12. Next, independent samples *t*-tests were conducted to examine if the relationships between dyad ratings differed by gender. Regarding child-caregiver dyads, the magnitude of male symptom discrepancies with the magnitude of female symptom discrepancies was not significant when rating internalizing symptoms ($t(151.53) = 0.1$, $p = 0.92$) or externalizing behaviors ($t(143.33) = 0.29$, $p = 0.77$). Regarding caregiver-teacher dyads, the magnitude of male symptom discrepancies with the magnitude of female symptom discrepancies was not significant when rating internalizing symptoms ($t(109.44) = 1.47$, $p = 0.14$) or externalizing behaviors ($t(112.85)$

= 1.05, p = 0.29). Finally, among child-teacher dyads, the magnitude of male symptom discrepancies with the magnitude of female symptom discrepancies was not significant when rating internalizing symptoms (t(111.41) = 0.08, p = 0.93) or externalizing behaviors (t(111.56) = 0.61, p = 0.54). Thus, the magnitude of discrepancy was unrelated to child gender regardless of reporting dyad across symptom cluster.

Effect of Race on Magnitudes of Discrepancy Across Dyads at age 12. To examine differences in race/ethnicity, the variable was dichotomously coded as "Black" and "Not Black." Independent samples *t*-tests were conducted to examine how child-caregiver discrepancy ratings varied as a function of race. Differences based on race were not observed among caregiver-child (t(107.13) = -0.98, p = 0.33) or teacher-caregiver (t(82.63) = -1.34, p = 0.18) dyads rating total symptoms. However, the magnitude of Black total symptom discrepancies with the magnitude of Not Black total symptom discrepancies was significant, (t(110.85) = -3.93, p < .001).On average, Black total symptom discrepancies (M = 0.62, SD = 0.42) were lower than Not Black total symptom discrepancies (M = 1.07, SD = 0.8).

Aim Two

Study timepoint was entered as the outcome variable while magnitude and direction were entered as predictor variables in the trajectory analyses to understand how caregiver-teacher discrepancies in rating regarding internalizing symptoms, externalizing behaviors, and total symptoms change over time. Magnitude and direction variables were entered to understand how interrater discrepancy changes over time. Data from four timepoints (e.g., caregiver-teacher ratings from child ages 10, 12, 14, and 16) were entered into a growth mixture model to determine whether groups of reporters followed similar level of disagreement over time.

After fitting an initial one-class model, two through ten-class growth mixture models were conducted for internalizing symptoms, externalizing behaviors, and total symptoms respectively and model fit was evaluated using multiple fit information criteria.

For internalizing symptoms (Table 6), initial analysis used entropy as a guiding fit index as prior literature indicates entropy values of >.80 are optimal for interpretation; the two- and three-class models had entropy values acceptable for further analysis. The three-class solution was selected given lower AIC, BIC, ssBIC, and CAIC values as compared to the two-class solution, as well as an adequate number of dyads in each class (Table 7).

Table 6. *Internalizing Symptom Growth Mixture Model Fit Indices*

	Loglike-lihood	AIC	BIC	ssBIC	CAIC	CLC	ICL-BIC	NEC	E
1	-946.31	1894.6	1898.34	1895.16	1899.34	1892.62	1898.34	1.00	1.00
2	-891.00	1786.01	1793.43	1787.09	1795.43	1815.36	1826.79	0.30	0.92
3	-881.68	1769.35	1780.49	1770.98	1783.49	1869.21	1886.35	0.82	0.84
4	-873.43	1754.85	1769.71	1757.02	1773.71	1934.58	1957.44	1.29	0.78
5	-873.43	1756.85	1775.42	1759.56	1780.42	2258.14	2286.71	3.51	0.48
6	-873.43	1748.85	1752.57	1749.39	1753.57	2326.02	2331.73	3.97	0.59
7	-873.43	1750.85	1758.28	1751.94	1760.28	2458.84	2470.27	4.88	0.69
8	-873.43	1752.85	1763.99	1754.48	1766.99	2594.69	2611.83	5.82	0.27
9	-860.06	1728.11	1742.97	1730.28	1746.97	2614.04	2636.89	5.18	0.06
10	-873.43	1756.85	1775.42	1759.56	1780.42	2786.05	2814.62	7.13	0.07

Table 7. *Number of Dyads per Internalizing Symptom Discrepancy Class*

Class	2	3	4	5
1	281	270	12	16
2	22	21	22	*
3		12	252	17
4			17	249
5				*

* No participants classified into that class

Class 1 is typified by caregiver-teacher dyads who have lower levels of disagreement across timepoints, while class 2 represents caregiver-teacher dyads with relatively consistent, high levels of disagreement. Though the third-class is smaller, it represents a distinct trajectory of caregiver-teacher dyads with high levels of disagreement at timepoint 1 who increase in level of agreement overtime (Figure 1). Trajectories are relatively consistent regardless of discrepancy direction (e.g., whether caregiver or teacher rated symptoms higher; Figure 2).

Figure 1. *Two- and Three-class Solutions for Internalizing Symptom Discrepancy*

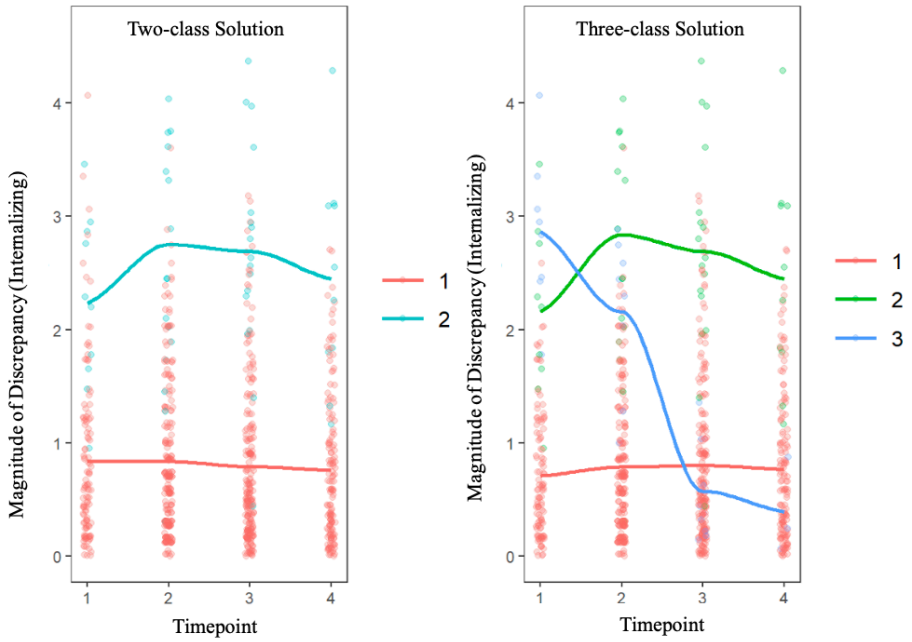
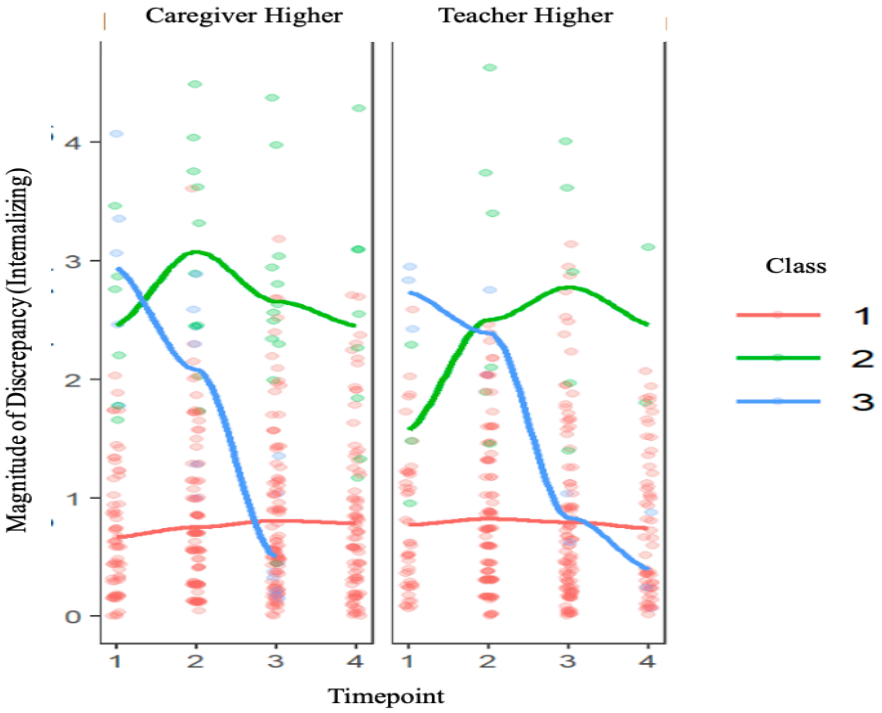


Figure 2. *Internalizing Symptom Trajectory Based on Direction of Discrepancy*



Regarding externalizing behaviors, entropy values of the two-class and three-class solutions indicated appropriate class homogeneity for further analysis (Table 8). While both solutions contained an adequate number of dyads in each class (Table 9), the three-class solution was selected given lower AIC, BIC, ssBIC, and CAIC values.

Table 8. *Externalizing Behaviors Growth Mixture Model Fit Indices*

	Loglike-lihood	AIC	BIC	ssBIC	CAIC	CLC	ICL-BIC	NEC	E
1	-838.93	1679.85	1683.57	1680.39	1684.57	1677.85	1683.57	1.00	1.00
2	-782.27	1568.54	1575.97	1569.63	1577.97	1623.12	1634.55	0.52	0.92
3	-750.74	1507.49	1518.63	1509.11	1521.63	1591.54	1608.68	0.52	0.84
4	-747.25	1502.50	1517.36	1504.67	1521.36	1694.32	1717.18	1.09	0.78
5	1756.85	1493.38	1511.95	1496.09	1516.95	1875.60	1904.17	2.02	0.48
6	-741.69	1485.38	1489.09	1485.92	1490.09	2041.82	2047.53	2.87	0.56

Table 9. *Number of Dyads per Externalizing Behavior Discrepancy Class*

Class	2	3	4	6
1	37	256	252	*
2	266	20	20	52
3		27	19	196
4			12	13
5				19
6				23

* No participants classified into that class

Within the three-class solution, class one represents dyads with consistently low levels of disagreement, class two contains dyads with increasing levels of disagreement over time, and class three represents dyads with generally decreasing levels of disagreement over time (Figure 3). Trajectory pattern is relatively consistent whether caregivers or teachers had higher behavior ratings (Figure 4).

Figure 3. *Two- and Three-class Solutions for Externalizing Symptom Discrepancy*

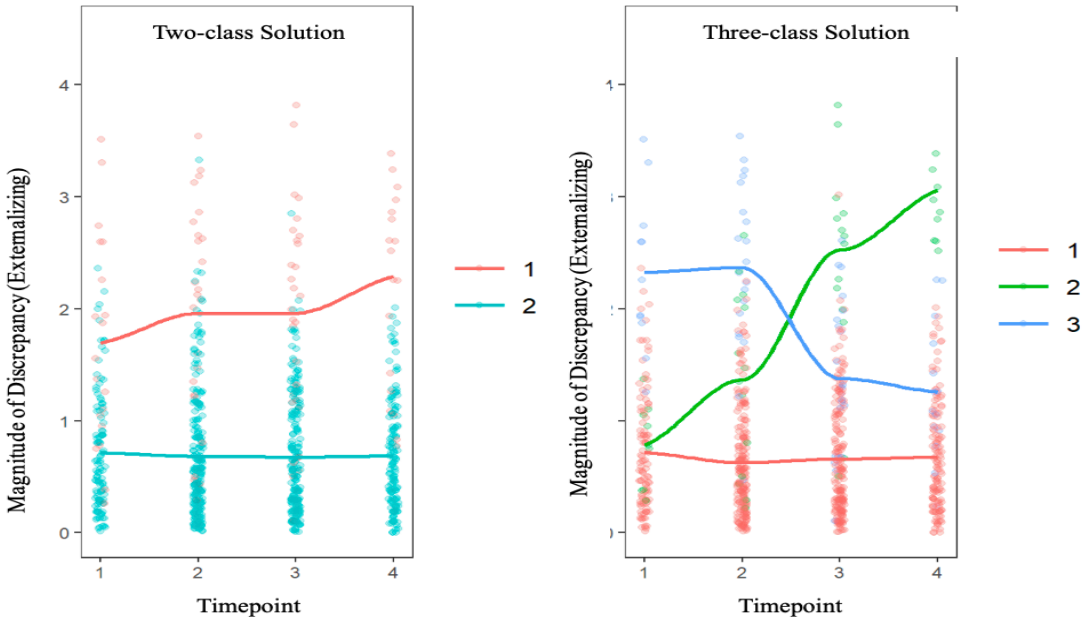
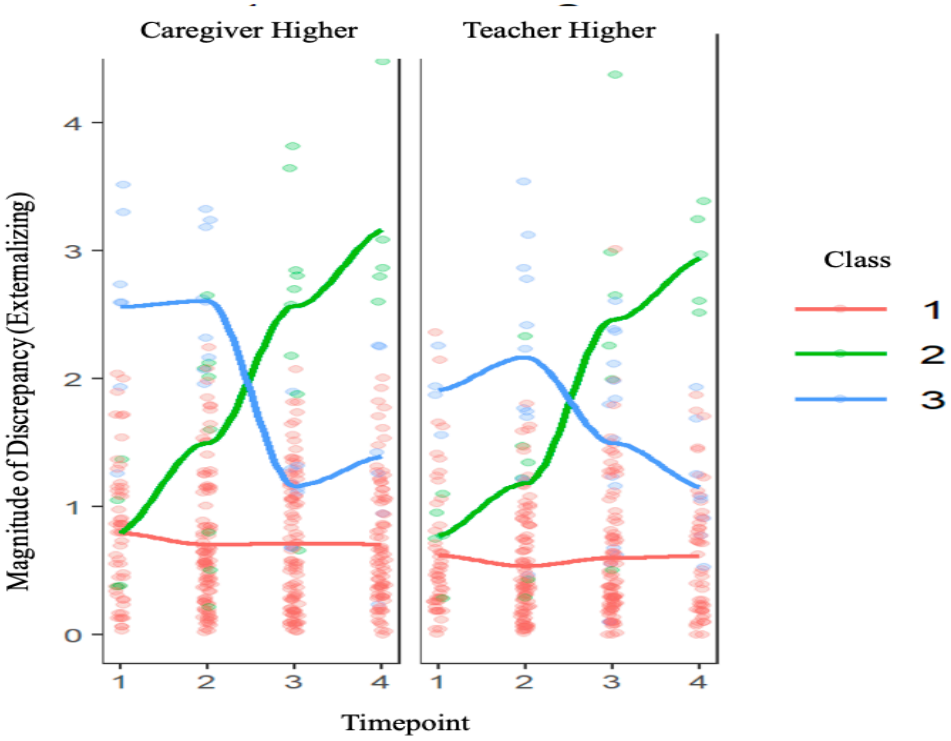


Figure 4. *Externalizing Behavior Trajectory Based on Direction of Discrepancy*



Finally, trajectories of caregiver-teacher discrepancy regarding total symptom ratings were analyzed. The two-, three- four-, and five-class solutions had adequate entropy for further analysis. While the five-class solution had more optimal fit indices when compared to other solutions, there were several challenges precluding the selection of this class (Table 10). Importantly, the addition of a fifth class into the model parses the prior solutions fourth class into smaller groups with less meaningful information, particularly once further divided based on direction of the discrepancy (Table 11). Thus, the four-class solution was selected.

Table 10. *Total Symptom Growth Mixture Model Fit Indices*

	Loglike-lihood	AIC	BIC	ssBIC	CAIC	CLC	ICL-BIC	NEC	E
1	-873.53	1749.07	1752.78	1749.61	1753.78	1747.07	1752.78	1.00	1.00
2	-815.43	1634.85	1642.28	1635.94	1644.28	1679.35	1690.78	0.42	0.88
3	-797.27	1600.54	1611.68	1602.16	1614.68	1675.19	1692.33	0.53	0.88
4	-783.61	1575.22	1590.08	1577.39	1594.08	1711.09	1733.95	0.79	0.83
5	-775.34	1560.68	1579.25	1563.39	1584.25	1740.14	1768.71	0.96	0.81
6	-773.29	1548.58	1552.29	1549.13	1553.29	1831.42	1837.14	1.42	0.54
7	-767.87	1550.58	1558.01	1551.67	1560.01	2114.42	2125.85	2.83	0.35
8	-767.87	1541.73	1552.87	1543.36	1555.87	1904.02	1921.16	1.74	0.45
9	-773.29	1554.58	1569.44	1556.75	1573.43	2396.71	2419.57	4.24	0.01
10	-773.29	1556.58	1575.15	1559.29	1580.15	2361.42	2389.99	4.06	0.16

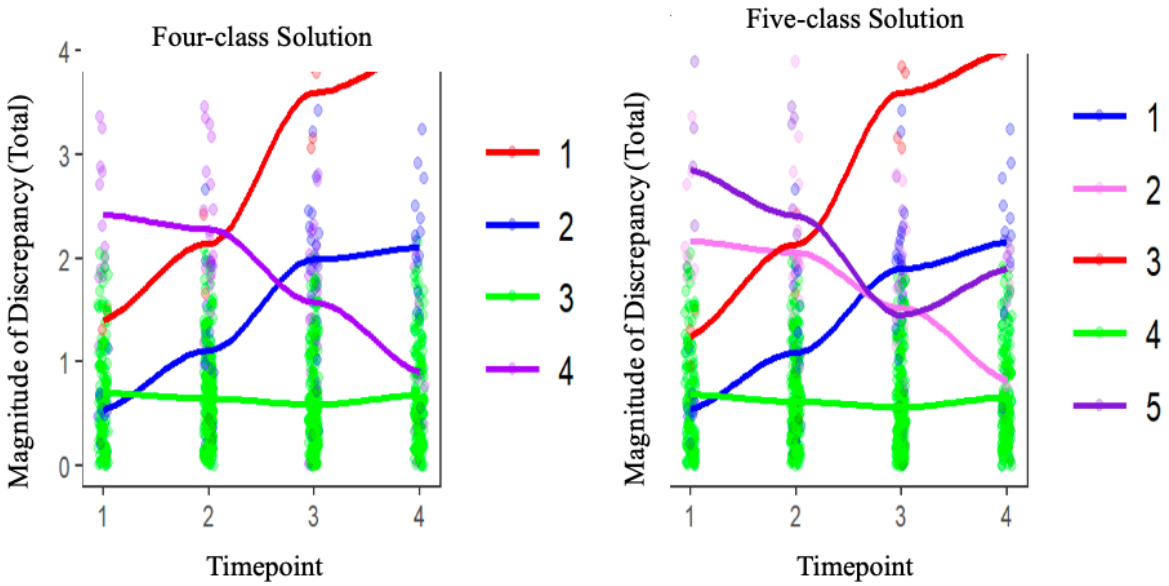
Table 11. *Number of Dyads per Total Symptom Discrepancy Class*

Class	2	3	4	5
1	273	262	6	27
2	30	8	26	26
3		33	241	7
4			30	232
5				11

* No participants classified into that class

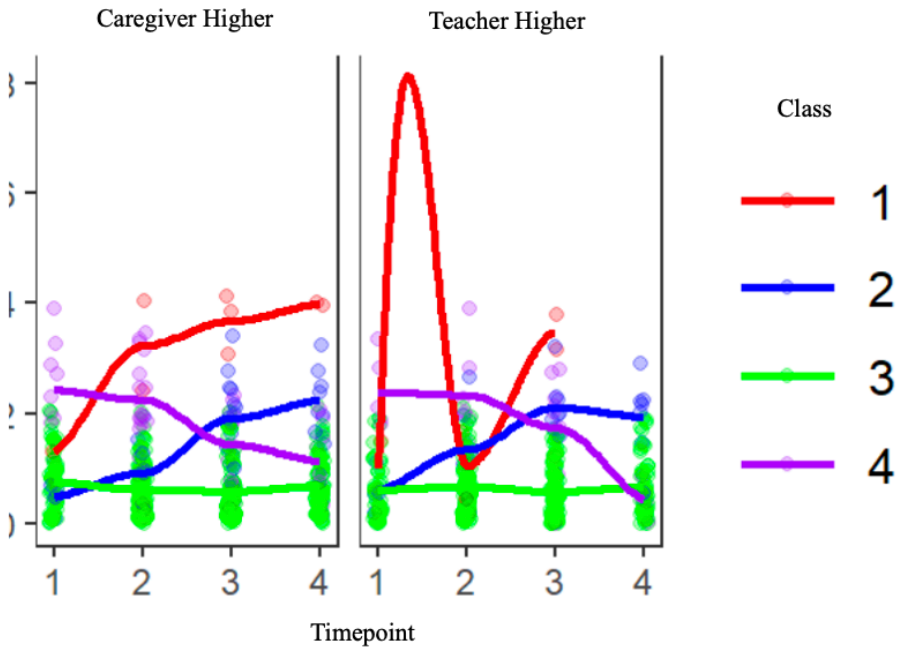
Within this solution, class one begins with moderate levels of caregiver-teacher discrepancy that increases over time. Class two is typified by dyads with low levels of disagreement that increases over time. Class three represents dyads with consistent low disagreement over time. Finally, class four contains dyads that begin with relatively high levels of disagreement the decreases over time (Figure 5).

Figure 5. *Four- and Five-class Solutions for Total Symptom Discrepancy*



However, these trajectories are markedly different when also examining discrepancy direction using loess curves (Figure 6). Specifically, when teachers report higher total symptomology, class one is typified by discrepancy that significantly increases between timepoint one and two before returning to baseline and steadily increasing to timepoint three. The remaining classes are relatively consistent regardless of discrepancy direction.

Figure 6. *Total Symptom Trajectory Based on Direction of Discrepancy*



Aim Three

After the appropriate class solution was selected, multiple regression analyses were conducted to determine the effect of caregiver-teacher discrepancy on child-reported mental health symptoms at age 18 when controlling for baseline ratings at age 12. Assumptions regarding normal distribution and multicollinearity were confirmed for internalizing symptoms, externalizing behaviors, and total symptom ratings. However, the assumption of homoscedasticity was violated across symptom cluster. As such, regression models more reliably predicted outcomes when symptoms are low as compared to when symptoms are high.

In contrast with hypotheses 3.1, 3.2, and 3.3, there is no significant relationship between class membership and distal child-rated internalizing outcomes (Table 12). While the overall model was significant ($F(3, 493) = 48.47$, $p < .01$) and accounted for 22.8% of the total observed variance, significance was largely driven by the positive predictor of child-rated internalizing

symptoms at age 12 (β = 0.50, p<0.01). There was not a significant change when adding class into the regression model (R^2 = 0.0004, p = .89). Specifically, trajectory one membership outcomes did not significantly differ from trajectory two membership outcomes (β = -0.01, p = 0.95) and trajectory three membership outcomes (β = 0.12, p = 0.64), nor did outcomes differ by membership between trajectories two and three (β = 0.13, p = 0.66).

Table 12. *Regression Statistics for Internalizing Symptom Trajectories*

	Predictor	β	β Error	*t*-value	*p*-value
Block One	(Intercept)	-0.01	0.04	-0.32	0.75
	Age 12	0.50	0.04	12.07	<0.01
Block Two	(Intercept)	-0.02	0.04	-0.36	0.72
	Age 12	0.50	0.04	11.69	<0.01
	Class 2*	-0.01	0.17	-0.06	0.95
	Class 3*	0.12	0.26	0.47	0.64

* Class one reference group
-- Class 2 to 3 comparison β = 0.13, p = 0.66

Regarding externalizing behaviors, a significant relationship between class membership and distal externalizing outcomes was observed (F(3, 493) = 39.93, p < 0.01; Table 13). Specifically, baseline child-rated mental health symptoms significantly positively predicted child-rated mental health at 18 (β = 0.41, p <0.01) and accounted for 18% of the total variance in child-rated externalizing behaviors at age 18. In addition, there was a significant change when adding class into the regression (R^2= 0.01, p = .033). While trajectory three membership outcomes did not differ compared to trajectory two outcomes (β = -0.15, p = 0.47) and trajectory one outcomes (β = 0.23, p =0.09), trajectory differences were seen when comparing trajectories one and two. Specifically, child-reported externalizing behaviors at age 18 are significantly higher for dyads included in class two (e.g., increasing disagreement over time; β = 0.23, p =0.03) as compared to class one (e.g., consistently low levels of disagreement).

Table 13. *Regression Statistics for Externalizing Behavior Trajectories*

	Predictor	β	β Error	*t*-value	*p*-value
Block One	(Intercept)	-0.08	0.04	-2.03	0.04
	Age 12	0.50	0.04	12.07	<0.01
Block Two	(Intercept)	-0.12	0.04	-2.89	<0.01
	Age 12	0.41	0.04	9.51	<0.01
	Class 2*	0.38	0.17	2.19	0.02
	Class 3*	0.23	0.14	1.66	0.09

* Class one reference group
-- Class 2 to 3 comparison β = -0.15, p = 0.48

Regarding child rated total symptoms at age 18, there are several differences between classes. The total model was significant (F(4, 492) = 47.98, p < 0.01; Table 14) and accounted for 28.1% of the total outcome variance. Baseline child rated total symptoms significantly predicted outcomes at 18 (β = 0.23, p < 0.01). In addition, there was a significant change when adding class into the regression (R^2 = 0.04, p < .01). In contrast to hypothesis 3.9, membership in trajectory one (e.g., moderate levels of disagreement that increase over time) significantly predicted lower level of child-rated total symptoms at age 18 as compared to trajectory 2 (β = 1.35, p < 0.01), trajectory 3 (β = 0.54, p = 0.04), and trajectory 4 (β = 0.77, p = 0.01). In contrast, membership in class two (e.g., mild disagreement that increases over time) predicted significantly higher child-rated total symptoms at age 18 as compared to trajectory 3 (β = -0.81, p < 0.01) and trajectory 4 (β = -0.58, p = 0.01). In contrast, outcomes did not significant differ by membership in trajectories three and four (β = -0.23, p = 0.10).

Table 14. *Regression Statistics for Total Symptom Trajectories*

	Predictor	β	β Error	*t*-value	*p*-value
Block One	(Intercept)	-0.02	0.04	-0.44	0.66
	Age 12	0.51	0.04	12.39	<0.01
Block Two	(Intercept)	-0.62	0.27	-2.29	0.02
	Age 12	0.48	0.04	11.66	<0.01
	Class 2*	1.35	0.31	4.32	<0.01
	Class 3*	0.54	0.28	1.97	0.04
	Class 4*	0.48	0.29	2.58	0.01

* Class one reference group

-- Class 2 to 3 comparison β = -0.81, p <0.01

-- Class 2 to 4 comparison β = -0.58, p <0.01

CHAPTER FIVE

DISCUSSION

The purpose of the current study was to quantify the level of disagreement between caregivers, teachers, and children when rating child mental health symptoms, to describe changes in disagreement between the caregiver-teacher dyad over adolescence, and to determine how the trajectory of disagreement affects child-reported mental health outcomes at age 18. This paper is a unique contribution to the literature as it not only captures changes in agreement over time, but also across symptom type (e.g., internalizing symptoms and externalizing behavior), whereas prior literature has typically examined disagreement at a single timepoint and/or for specific disorders (e.g., depression). Thus, the current study was better able to capture the complexity of multi-informant ratings in a more holistic way as compared to more traditional methods. Further, this study was conducted using data collected from the chronically understudied population of children and families with prior and/or ongoing child welfare involvement.

Multi-informant Agreement and Related Child Factors

The 12-year-old dataset provided the greatest opportunity to capture inter-rater agreement since it provided data for children, caregivers, and teachers (i.e., caregiver-child, child-teacher, and teacher-caregiver dyads). Consistent with the prior literature (De Los Reyes, 2015), ratings ranged from low-to-moderate agreement across caregiver-child (r = .32-.43), child-teacher (r = .25-.37), and teacher-caregiver dyads (r = .24-.33). However, the child-related factor of gender did not moderate interrater discrepancies regarding externalizing behaviors or internalizing symptoms at child age 12. This is notably different than the extant literature. For example,

McWey et al. (2018) found that foster caregivers and youth are less likely to agree on mental health ratings when the adolescent is male. In addition, race was related to magnitude of discrepancy among child-teacher dyads. However, where prior literature suggests that systemic oppression (Pactor & Garcia Coll, 2009) and teacher racism (Kang & Harvey, 2020) contribute to significantly more discrepant reports, the present study indicates that Black children and teachers have smaller magnitudes of discrepancy as compared to non-Black children.

Taken together, these results underscore the importance of multi-informant reporting and indicate the importance of considering history of child welfare involvement as a unique context in multi-informant assessment. Specifically, biological parents and foster caregivers represent distinct reporters and literature related to one population may not apply to another. Future studies should examine the extent to which additional caregiving dyads in the child welfare system, like kinship caregivers, agree with child and teacher reports given that the extant literature indicates that kinship care is associated with several positive outcomes as compared to traditional foster care (Xu & Bright, 2018). Further, clinicians and child welfare workers should consider that multi-informant reporting may be as reliable for boys as for girls, and as for Black children as compared to non-Black children, regardless of symptom cluster and thus should not discount caregiver and Black boys reporting of internalizing symptoms within this population.

Growth Mixture Models and Associated Discrepancies

In addition, growth mixture modeling indicated several significant findings. As described, growth mixture modeling identified three trajectories of dyadic discrepancies for internalizing symptom reports: 1) reporters that have consistently low levels of disagreements across timepoints, 2) reporters that have consistently high levels of disagreement, and 3)

reporters that begin with high levels of disagreement that decreases (e.g., better agreement) over time. Trajectories were consistent regardless of whether caregivers or teachers rated symptoms more highly. Class membership did not significantly predict child-reported internalizing symptoms at age 18.

There were also three identified trajectories among caregiver-teacher dyads reporting on externalizing behaviors. Specifically, regardless of whether teacher or caregiver rated symptoms more highly, trajectories included 1) reporters that have consistently low levels of disagreements across timepoints, 2) reporters that have increasing disagreement over time (low to moderate), and 3) reporters that have decreasing disagreement over time (moderate to low). Dyads associated with increasing levels of discrepancy over time were associated with higher child-reported externalizing behaviors at age 18.

Finally, four trajectories were identified for dyads rating total mental health symptoms across adolescence. Trajectories included 1) reporters with moderate levels of discrepancy at baseline which increase across adolescence (moderate to high), 2) reporters with low levels of disagreement at baseline that increases over time (low to moderate), 3) reporters with consistently low disagreement, and 4) reporters beginning with high disagreement that decreases over time (moderate to low). The only trajectory that differed based on which reporter rated symptoms as more severe was trajectory one. Specifically, when teachers reported a higher symptom burden than caregivers, the trajectory was associated with a spike in disagreement between ages 10 and 12. In addition, membership in trajectory two (low to moderate disagreement) predicted higher child-reported total symptoms at age 18, whereas membership in trajectory 1 (moderate to high disagreement) predicted lower symptoms.

Taken together, there are several similarities in trajectories across symptom cluster. First, regardless of symptom cluster, the majority of children were included in trajectories typified by consistently low levels of caregiver-teacher disagreement. This suggests that, in general, positive mesosystems functioning among children with prior and/or ongoing child welfare involvement, as interrater reliability increases with relational closeness among reporters (Van Dulmen & Egeland, 2011). This is particularly important as strained relationships between school and child welfare stakeholders is associated with increased stress and academic challenges (Moyer & Goldberg, 2020). Importantly, however, consistently low caregiver-teacher disagreement was not significantly associated with levels of child-reported symptomology at age 18 when compared to other trajectories which included higher levels of discrepancy. It may be that caregiver-teacher dyads were not accurately capturing youth internalizing symptoms, particularly as internalizing symptoms are often the most difficult for outside observers to assess accurately (McWey et al., 2014) and youth who have experienced abuse and/or neglect are more likely to experience internalizing symptoms. It may also be that functioning mesosystems are not enough to bolster the mental health of youth with prior and/or ongoing child welfare involvement experience given the higher-than-average rates of complex trauma that this population experiences.

In addition, a number of trajectories were associated with discrepancy changes over time. This is consistent with the extant literature that suggests interrater discrepancies are not static. The two general changes over time included dyads with decreasing discrepancy (e.g., internalizing trajectory three, total symptom trajectory four) and those with increasing discrepancy over time (e.g., externalizing trajectory three, total symptom trajectory one, total symptom trajectory two). Notably, only membership in trajectories with increasing discrepancy

over time predicted child-rated mental health outcomes at age 18. However, outcomes differed depending on the magnitude of discrepancy over time.

Specifically, children reported greater mental health symptom burden at age 18 when caregivers and teachers had increasing disagreement ending in the moderate range (externalizing trajectory two and total symptom trajectory two). There are several possible explanations for these findings. First, it may be symptoms were more apparent in one setting, as thus dyads were observing differing levels of mental health symptoms (De Los Reyes et al., 2015) and were not communicating regarding the adolescent's functioning across setting. Within this context, youth may not be receiving adequate wraparound social support, particularly as youth involved child welfare require no less than three varying sources of support to experience the psychosocial benefits (Perry, 2006). Further, it may be that conflict between caregivers and teachers is noticed by adolescents and contributes to increased externalizing challenges.

In contrast, membership in total symptom trajectory one, which was associated with increasing disagreement ending in the high range, predicted lower levels of child-rated mental health symptoms at 18. This suggests that there may be a level of discrepancy associated with improved outcomes. It may be that youth experiencing increasing such notable caregiver-teacher disagreement are provided with the appropriate services (e.g., psychotherapy, pharmacological intervention) and are thus better able to cope with their mental health symptoms. It may also be that the child experiences psychosocial support and subsequently benefits from an attuned adult who understands and agrees with their own interpretation of their behavior. Future studies should examine service utilization as a moderator of the relationship between caregiver-teacher disagreement trajectory and child-rated mental health outcomes.

Finally, the results indicated a significant difference among dyads rating total mental health symptoms based on the direction of the discrepancy (e.g., caregiver or teacher rating symptoms as more clinically prominent). Specifically, discrepancies increased significantly between timepoint one and two before returning to baseline and steadily increasing to timepoint three when teachers reported more prominent mental health symptoms. Notably, this spike occurred between ages 10 and 12. It may be that the executive functioning-related demands of school increase at this time and contribute to increased externalizing symptoms (Jacobson et al., 2010), while puberty-related changes and social stress contribute to increased internalizing symptoms during this developmental period (Rapee et al., 2019).

Limitations

There are several limitations to consider. As data were collected as part of a larger multi-site longitudinal study, there were several methodological challenges. First, TRFs were not collected at the Northwest site at child aged 10. Relatedly, YSR data were not solicited at ages 14 and 16. Thus, it is not possible to control for changes in child-rated mental health symptoms within the growth mixture model. In addition, the child ages represent a window in which data was collected and thus include varying chronological ages (LONGSCAN Manual). Further, the guardian completing the CBCL was not specified at each timepoint; thus, it was not possible to control for changes in caregivers across trajectories. As such, discrepancies may reflect caregiver changes rather than symptom changes or interrater disagreement. Therefore, placement instability may account for some of the differences based on class membership in child-reported mental health outcomes at age 18. Future studies should specify whether adoptive or biological parent is completing rating measures, as well as any caregiver changes across timepoints.

Relatedly, ongoing child welfare involvement was not captured in the LONGSCAN study. While the sample was restricted to only include children who had contact with the child welfare system, it is not possible to determine the extent to which families received child abuse and neglect prevention services and/or foster care services with the current data. Future studies should evaluate the extent to which system involvement, as well as lifetime experiences of abuse and neglect, affect reporter discrepancies and their changes over time.

As previously noted, a significant proportion of the sample was missing data regarding gender and race/ethnicity. As such, it was not possible to include demographic variables into regression models. Relatedly, the regression models were heteroscedastic, indicating that additional variables are needed to account for the observed variance in child-rated mental health symptoms at age 18. Future studies should examine additional predictors of child-rated mental health above and beyond caregiver-teacher discrepancy trajectory. In particular, prior literature suggests that individual primary developmental processes are similar across race/ethnicity; rather, the interaction between the child from a historically minoritized population and systems of oppression and experiences of racism influence development (see Garcia Coll et al., 1996). This is particularly relevant among youth with prior and ongoing child welfare involvement given the child welfare systems' history of systemic oppression. Thus, future studies should include more nuanced measures of social identity (e.g., experiences of racism, experiences of sexism), as well as other systemic factors (e.g., community violence; Fowler et al., 2009) known to increase mental health challenges among historically minoritized populations.

Implications

Despite these challenges, the current study makes notable contributions to the multi-informant literature with several practical implications. From a research perspective, the present findings undergird the growing consensus that multi-informant discrepancies capture meaningfully different information rather than instrument error. As such, researchers investigating child mental health should prioritize collecting many and varied perspectives, particularly among the child welfare-involved population. In addition, the discrepancy trajectories significantly differed based on direction when rating total youth symptomology. Thus, it is important that researchers capture not only the magnitude of the discrepancy, but the direction of the discrepancy as well in order to better describe the nuance related to interrater differences.

In addition, there is nuance regarding the role of interrater discrepancy on child mental health functioning. Indeed, discrepancies do not have a universally positive or negative influence on the child's mental health functioning. Rather, the outcomes are dependent on factors driving the discrepancy. First, increasing discrepancies may indicate that the adolescent's social support system is not functioning appropriately, which suggests that increasing communication and support should be a priority in casework practice. Next, interrater disagreement may suggest that a caregiver or teacher is not accurately identifying a child's needs, which in turn may contribute to missing key therapeutic intervention windows and downstream mental health effects. Thus, increasing levels of caregiver-teacher discrepancy should act as a warning sign for clinicians and caseworkers addressing child mental health needs.

REFERENCE LIST

Achenbach, T. M. (2009). *The Achenbach System of Empirically Based Assessemnt (ASEBA): Development, Findings, Theory, and Applications*. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.

Achenbach, T. M., Becker, A., Dopfner, M., et al. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions. *Journal of Child Psychology and Psychiatry, 49*, 251–275. https://doi.org/10.1111/j.1469-7610.2007.01867.x

Achenbach, T. M., Dumenci, L., & Rescorla, L. A. (2002). Ten-year comparisons of problems and competencies for national samples of youth: Self, parent, and teacher reports. *Journal of Emotional and Behavioral Disorders*, *10*(4), 194–203. https://doi.org/10.1177/10634266020100040101

Achenbach, T.M., & Rescorla, L.A. (2001). Manual for the ASEBA school-age forms & profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Achenbach, T., Ivananova, M., & Rescorla, L. (2019). Multi-informant assessment of psychopathology from preschool through old age. In *The Cambridge Handbook of Clinical Assessment and Diagnosis* (pp. 123–134). Cambridge University Press. https://doi.org/10.1017/9781108235433.011

Alvidrez, Snowden, L. R., & Kaiser, D. M. (2008). The Experience of stigma among Black mental health consumers. *Journal of Health Care for the Poor and Underserved*, 19(3), 874–893. https://doi.org/10.1353/hpu.0.0058

American Psychological Association. (2013). *Standards for educational and psychological testing*. American Educational Research Association.

an der Ende, Verhulst, F., & Tiemeier, H. (2012). Agreement of informants on emotional and behavioral problems from childhood to adulthood. *Psychological Assessment*, *24*(2), 293–300. https://doi.org/10.1037/a0025500

Augenstein, Visser, K. F., Gallagher, K., De Los Reyes, A., D'Angelo, E. J., & Nock, M. K. (2022). Multi-informant reports of depressive symptoms and suicidal ideation among adolescent inpatients. *Suicide & Life-Threatening Behavior,* 52(1), 99–109. https://doi.org/10.1111/sltb.12803

Bagby, M., & Solomon-Krakus, S. (2019). Writing a psychological report using evidence-based psychological assessment methods. In *The Cambridge Handbook of Clinical Assessment and Diagnosis* (pp. 101–110). Cambridge University Press. https://doi.org/10.1017/9781108235433.009

Barboza, G. E., & Dominguez, S. (2017). Longitudinal growth of post-traumatic stress and depressive symptoms following a child maltreatment allegation: An examination of violence exposure, family risk and placement type. *Children and Youth Services Review*, *81*, 368–378. https://doi.org/10.1016/j.childyouth.2017.08.029

Becker-Haimes, E. M., Jensen-Doss, A., Birmaher, B., Kendall, P. C., & Ginsburg, G. S. (2018). Parent–youth informant disagreement: Implications for youth anxiety treatment. *Clinical Child Psychology and Psychiatry*, 23(1), 42-56. https://doi.org/10.1177/1359104516689586

Berrick, J. (2006). Neighborhood-based foster care: A critical examination of location-based placement criteria. *The Social Services Review,* 80(4), 569-583. https://doi.org/10.1086/507933

Blakemore, SJ., & Mills, K. L. (2014). Is adolescence a sensitive period for sociocultural processing? *Annual Review of Psychology*, *65*(1), 187–207. https://doi.org/10.1146/annurev-psych-010213-115202

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis,* 12, 3-27. https://doi.org/10.1093/pan/mph001

Branje, S. (2018). Development of parent–adolescent relationships: Conflict interactions as a mechanism of change. *Child Development Perspectives*, 12(3), 171–176. https://doi.org/10.1111/cdep.12278

Brinkman, Epstein, J. N., Auinger, P., Tamm, L., & Froehlich, T. E. (2014). Association of attention-deficit/hyperactivity disorder and conduct disorder with early tobacco and alcohol use. *Drug and Alcohol Dependence*, *147*, 183–189. https://doi.org/10.1016/j.drugalcdep.2014.11.018

Bronsard, G., Alessandrini, M., Fond, G., Loundou, A., Auquier, P., Tordjman, S., & Boyer, L. (2016). The prevalence of mental disorders among children and adolescents in the child welfare system: A systematic review and meta-analysis. *Medicine (Baltimore)*, *95*(7), e2622–e2622. https://doi.org/10.1097/MD.0000000000002622

Burk, W. J., & Laursen, B. (2010). Mother and adolescent reports of associations between child behavior problems and mother-child relationship qualities: Separating shared variance from individual variance. *Journal of Abnormal Child Psychology,* 38(5), 657–667. https://doi.org/10.1007/s10802-010-9396-z

Butcher, J. N. (2006). Assessment in clinical psychology: A perspective on the past, present challenges, and future prospects. *Clinical Psychology: Science and Practice*, 13(3), 205–209. https://doi.org/10.1111/j.1468-2850.2006.00025.x

Capacity Building Center for States. (2017). Collaborating to collect and use data to inform quality contacts. Washington, DC: U.S. Department of Health and Human Services, Children's Bureau. Retrieved from https://capacity.childwelfare.gov/states/resources/collect-use-data-quality-contacts

Capacity Building Center for States. (2017). Decision-making in child welfare for improved safety outcomes. Retrieved from: https://capacity.childwelfare.gov/states/resources/decision-making-improved-safety-outcomes/

Chibnall, S., Dutch, N., Jones-Harden, B., Brown, A., Gourdine, R., Boone, A., & Snyder, S. (2003). Children of color in the child welfare system: Perspectives from the child welfare community. *Administration for Children and Families.*

Chor, McClelland, G. M., Weiner, D. A., Jordan, N., & Lyons, J. S. (2014). Out-of-home placement decision-making and outcomes in child welfare: A longitudinal study. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(1), 70–86. https://doi.org/10.1007/s10488-014-0545-5.

Clark, S., & Muthén, B. (2009). Relating latent class analysis results to variables not included in the analysis. Retrieved from http://www.statmodel.com/download/relatinglca.pdf

Cohen. (1977). *Statistical power analysis for the behavioral sciences* (Revised edition.). Academic Press.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin,* 136(6), 1092–1122. https://doi.org/10.1037/a0021212

Darlington, Y., Feeney, J. A., & Rixon, K. (2005). Interagency collaboration between child protection and mental health services: Practices, attitudes and barriers. *Child Abuse & Neglect*, 29(10), 1085–1098. https://doi.org/10.1016/j.chiabu.2005.04.005

De Los Reyes, A., Augenstein, T., Wang, M., Thomas, S., Drabick, D., Burgers, D., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*(4), 858–900. https://doi.org/10.1037/a0038498

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin, 131*(4), 483–509. https://doi.org/10.1037/0033-2909.131.4.483

De Los Reyes, Ohannessian, C. M., & Racz, S. J. (2019). Discrepancies between adolescent and parent reports about family relationships. *Child Development Perspectives, 13(*1), 53–58. https://doi.org/10.1111/cdep.12306

Deighton, J., Croudace, T., Fonagy, P., Brown, J., Patalay, P., & Wolpert, M. (2014). Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: a review of child self-report measures. *Child and Adolescent Psychiatry and Mental Health,* 8(1), 14–14. https://doi.org/10.1186/1753-2000-8-14

Dettlaff, A. J., & Boyd, R. (2020). Racial disproportionality and disparities in the child welfare system: Why do they exist, and what can be done to address them? *The Annals of the American Academy of Political and Social Science,* 692(1), 253–274. https://doi.org/10.1177/0002716220980329

Ferdinand, R.F., van der Ende, J., Verhulst, F.C. (2004). Parent-adolescent disagreement regarding psychopathology in adolescents from the general population as a risk factor for adverse outcome. *Journal of Abnormal Psychology*, 113(2), 198-206. https.doi.org/10.1037/0021-843X.113.2.198

Fluke, J., Jones Harden, B., Jenkins, M., & Ruehrdanz, A. (2011). Research synthesis on child welfare disproportionality and disparities. Papers from a Research Symposium Convened by the Center for the Study of Social Policy and The Annie E. Casey Foundation on behalf of The Alliance for Racial Equity in Child Welfare. Retrieved from: https://casala.org/wp-content/uploads/2015/12/Disparities-and-Disproportionality-in-Child-Welfare_An-Analysis-of-the-Research-December-2011-1.pdf#page=11

Font, Sattler, K. M. ., & Gershoff, E. T. (2018). Measurement and correlates of foster care placement moves. *Children and Youth Services Review*, 91, 248–258. https://doi.org/10.1016/j.childyouth.2018.06.019

Fowler, P.J., Tompsett, C.J., Braciszewski, J.M., Jacques-Tiura, A.J., & Bates, B.B. (2009). Community violence: A meta-analysis on the effect of exposure and mental health outcomes of children and adolescents. *Development and Psychopathology*, 21(1), 227-259. https://doi.org/10.1017/S0954579409000145

Frankfurt, S., Frazier, P., Syed, M., & Jung, K. R. (2016b). Using group-based trajectory and growth mixture modeling to identify classes of change trajectories. *The Counseling Psychologist*, *44*(5), 622-660. https://doi.org/10.1177/0011000016658097

Garcia Coll, C., Lamberty, G., Jenkins, R., Pipes McAdoo, H., Crnic, K., Wasik, B.H., Vázquez Garcia, H. (1996). An integrative model for the study of developmental competencies in minority children. *Child Development,* 67(5), 1891-1914. PMID: 9022222

Gomez, R., Vance, A., & Gomez, R. M. (2014). Analysis of the convergent and discriminant validity of the CBCL, TRF, and YSR in a clinic-referred sample. *Journal of Abnormal Child Psychology*, *42*(8), 1413–1425. https://doi.org/10.1007/s10802-014-9879-4

Goolsby, Rich, B. A., Hinnant, B., Habayeb, S., Berghorst, L., De Los Reyes, A., & Alvord, M. K. (2017). Parent–child informant discrepancy is associated with poorer treatment outcome. *Journal of Child and Family Studies*, *27*(4), 1228–1241. https://doi.org/10.1007/s10826-017-0946-7

Grigorenko, E. L., Geiser, C., Slobodskaya, H. R., & Francis, D. J. (2010). Cross-informant symptoms from CBCL, TRF, and YSR: Trait and method variance in a normative sample of Russian youths. *Psychological Assessment, 22*, 893–911. https://doi.org/10.1037/a0020703

Gypen, L., Vanderfaeillie, J., De Maeyer, S., Belenger, L., & Van Holen, F. (2017). Outcomes of children who grew up in foster care: Systematic-review. *Children and Youth Services Review*, *76*, 74–83. https://doi.org/10.1016/j.childyouth.2017.02.035

Hennig, T., Schramm, S., & Linderkamp, F. (2017). Cross-informant disagreement on behavioral symptoms in adolescent attention-deficit/hyperactivity disorder and its impact on treatment effects. European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment, 34(2), 79–86. https://doi.org/10.1027/1015-5759/a000446

Heflinger, C., Simpkins, C., & Combs-Orme, T. (2000). Using the CBCL to determine the clinical status of children in state custody. *Children and Youth Services Review*, *22*(1), 55–73. https://doi.org/10.1016/S0190-7409(99)00073-0

Hoffman, & Chu, B. C. (2015). Target problem (mis) matching: Predictors and consequences of parent–youth agreement in a sample of anxious youth. *Journal of Anxiety Disorders*, *31*, 11–19. https://doi.org/10.1016/j.janxdis.2014.12.015

Hong, J. S., Algood, C. L., Chiu, Y.-L., & Lee, S. A.-P. (2011). An ecological understanding of kinship foster care in the United States. *Journal of Child and Family Studies*, *20*(6), 863–872. https://doi.org/10.1007/s10826-011-9454-3

Hunsley, J., & Mash, E. J. (2020). The role of assessment in evidence-based practice. In M. M. Antony & D. H. Barlow (Eds.), *Handbook of assessment and treatment planning for psychological disorders* (pp. 3–23). The Guilford Press.

Hunsley, J., & Mash, E. (2005). Evidence-based assessment of child and adolescent disorders: Issues & challenges. *Journal of Clinical Child and Adolescent Psychology,* 34, 362-370. https://doi.org/10.1207/s15374424jccp3403_1

Jackson, & King, A. R. (2004). Gender differences in the effects of oppositional behavior on teacher ratings of ADHD Symptoms. *Journal of Abnormal Child Psychology*, *32*(2), 215–224. https://doi.org/10.1023/B:JACP.0000019772.71251.ff

Jacobson, L. A., Williford, A. P., & Pianta, R. C. (2011). The role of executive function in children's competent adjustment to middle school. *Child Neuropsychology,* 17(3), 255–280. https://doi.org/10.1080/09297049.2010.535654

Jensen-Doss, A., Haimes, E. M. B., Smith, A. M., Lyon, A. R., Lewis, C. C., Stanick, C. F., & Hawley, K. M. (2018). Monitoring treatment progress and providing feedback is viewed favorably but rarely used in practice. *Administration and Policy in Mental Health and Mental Health Services Research*, *45*(1), 48–61. https://doi.org/10.1007/s10488-016-0763-0

Jensen-Doss, A., & Hawley, K. M. (2010). Understanding barriers to evidence-based assessment: Clinician attitudes toward standardized assessment tools. *Journal of Clinical Child & Adolescent Psychology,* 39(6), 885-896.

Kang, & Harvey, E. A. (2019). Racial differences between Black parents' and white teachers' perceptions of Attention-Deficit/Hyperactivity Disorder behavior. *Journal of Abnormal Child Psychology, 48*(5), 661–672. https://doi.org/10.1007/s10802-019-00600-y

Keller, Salazar, A.M., Courtney, M.E. (2010). Prevalence and timing of diagnosable mental health, alcohol, and substance use problems among older adolescents in the child welfare system. *Child Youth Services Review,* 32(4), 626–634. https://doi.org/10.1016/j.childyouth.2009.12.010

Kessler, R., Avenevoli, S., Costello, E., Georgiades, K., Green, J., Gruber, M., He, J., Koretz, D., McLaughlin, K., Petukhova, M., Sampson, N., Zaslavsky, A., & Merikangas, K. (2012). Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders in the national comorbidity survey replication adolescent supplement. *Archives of General Psychiatry*, *69*(4), 372–380. https://doi.org/10.1001/archgenpsychiatry.2011.160

Kisiel, C., Fehrenbach, T., Small, L., & Lyons, J.S. (2009). Assessment of complex trauma exposure, responses, and service needs among children and adolescents in child welfare. *Journal of Child & Adolescent Trauma, 2*, 143-160. https://doi.org/10.1080/19361520903120467

Lau, A. S., Garland, A. F., Yeh, M., McCabe, K. M., Wood, P. A., & Hough, R. L. (2004). Race/ethnicity and inter-informant agreement in assessing adolescent psychopathology. *Journal of Emotional and Behavioral Disorders, 12*, 145-156. https://doi.org/10.1177/10634266040120030201

Lawson, Nissley-Tsiopinis, J., Nahmias, A., McConaughy, S. H., & Eiraldi, R. (2017). Do parent and teacher report of ADHD symptoms in children differ by SES and racial status? *Journal of Psychopathology and Behavioral Assessment, 39*(3), 426–440. https://doi.org/10.1007/s10862-017-9591-0

Mack. (2021). The white supremacy hydra: How the Family First Prevention Services Act reifies pathology, control, and punishment in the family regulation system. *Columbia Journal of Race and Law*, 11(3). https://doi.org/10.52214/cjrl.v11i3.87514

Makol, B., De Los Reyes, A., Garrido, E., Harlaar, N., & Taussig, H. (2021). Assessing the mental health of maltreated youth with child welfare involvement using multi-informant reports. *Child Psychiatry and Human Development*, *52*(1), 49–62. https://doi.org/10.1007/s10578-020-00985-8

Martel, M., Markon, K., & Smith, G. (2017). Research review: Multi-informant integration in child and adolescent psychopathology diagnosis. *Journal of Child Psychology and Psychiatry*, *58*(2), 116–128. https://doi.org/10.1111/jcpp.12611

Martin, DeMarni Cromer, L., Freyd, J. (2010) Teachers' beliefs about maltreatment effects on student learning and classroom behavior. *Journal of Child & Adolescent Trauma, 3*(4), 245-254.

Mash, E. J., & Hunsley, J. (2005). Evidence-based assessment of child and adolescent disorders: Issues and challenges. *Journal of Clinical Child and Adolescent Psychology, 34*(3), 362–379. https://doi.org/10.1207/s15374424jccp3403_1

McWey, L., Cui, M., Cooper, A., & Ledermann, T. (2018). Caregiver–adolescent disagreement on the mental health of youth in foster care: The moderating role of the caregiver relationship. *Child Maltreatment*, *23*(3), 294–302. https://doi.org/10.1177/1077559518769375

McWey, L., Cui, M., & Holtrop, K. (2015). Caregiver–adolescent dyadic disagreement and problem behaviors of adolescents in the child welfare system. *Journal of Research on Adolescence*, *25*(1), 44–50. https://doi.org/10.1111/jora.12105

Medicine, I. of., Populations, B. on the H. of S., & Committee on Psychological Testing, I. V. T. (2015). *Psychological Testing in the Service of Disability Determination.* National Academies Press.

Merikangas, K. R., He, J., Burstein, M., Swendsen, J., Avenevoli, S., Case, B., Georgiades, K., Heaton, L., Swanson, S., & Olfson, M. (2011). Service utilization for lifetime mental disorders in U.S. adolescents: Results of the national comorbidity survey–adolescent supplement (NCS-A). *Journal of the American Academy of Child and Adolescent Psychiatry*, 50(1), 32–45. https://doi.org/10.1016/j.jaac.2010.10.006

Merkel-Holguin, L., Drury, I., Gibley-Reed, C., Lara, A., Jihad, M., Grint, K., & Marlowe, K. (2022). Structures of oppression in the U.S. child welfare system: Reflections on administrative barriers to equity. *Societies,* 12, 26. https://doi.org/10.3390/soc12010026

Moyer, A. M., & Goldberg, A. E. (2020). Foster youth's educational challenges and supports: Perspectives of teachers, foster parents, and former foster youth. *Child & Adolescent Social Work Journal*, *37*(2), 123–136. https://doi.org/10.1007/s10560-019-00640-9

Nylund-Gibson, K., Grimm, R.P., & Masyn, K.E. (2019) Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(6), 967-985, https://doi.org/10.1080/10705511.2019.1590146

O'Connell, M. E., Boat, T. F., & Warner, K. E. (2009). *Preventing mental, emotional, and behavioral disorders among young people progress and possibilities.* National Academies Press.

O'Higgins, A., Sebba, J., & Gardner, F. (2017). What are the factors associated with educational achievement for children in kinship or foster care: A systematic review. *Children and Youth Services Review*, *79*, 198–220. https://doi.org/10.1016/j.childyouth.2017.06.004

Parker, E., Parker, E., Jacobson, J., Jacobson, J., Pullmann, M., Pullmann, M., Kerns, S., & Kerns, S. (2019). Identifying psychosocial problems among children and youth in the child welfare system using the PSC-17: Exploring convergent and discriminant validity with multiple informants. *Child Psychiatry and Human Development*, *50*(1), 108–120. https://doi.org/10.1007/s10578-018-0824-6

Pecora, P., Jensen, P., Romanelli, L., Jackson, L., & Ortiz, A. (2009). Mental health services for children placed in foster care: An overview of current challenges. *Child Welfare*, *88*(1), 5–26. PMCID: PMC3061347

Perry, B. (2006). Understanding social network disruption: The case of youth in foster care. *Social Problems*, 53(3), 371-391. https://doi.org/10.1525/sp.2006.53.3.371

Ram, & Grimm K. (2007). Using simple and complex growth models to articulate developmental change: Matching theory to method. *International Journal of Behavioral Development,* 31, 303–316. https://doi.org/10.1177/0165025407077751

Ram, & Grimm, K. J. (2009). Methods and Measures: Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development*, *33*(6), 565–576. https://doi.org/10.1177/0165025409343765

Ram, N. & Grimm, K.J. (2009). Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavioral Development,* 33, 565-576. https://doi.org/10.1177/0165025409343765

Rapee, R., Oar, E., Johnco, C., Forbes, M., Fardouly, J., Magson, N., & Richardson, C. (2019). Adolescent development and risk for the onset of social-emotional disorders: a review and conceptual model. *Behaviour Research and Therapy, 123*, 103501. https://doi.org/10.1016/j.brat.2019.103501

Rescorla, Achenbach, T. M., Ivanova, M. Y., Harder, V. S., Otten, L., Bilenberg, N., Bjarnadottir, G., Capron, C., De Pauw, S. S. W., Dias, P., Dobrean, A., Döpfner, M., Duyme, M., Eapen, V., Erol, N., Esmaeili, E. M., Ezpeleta, L., Frigerio, A., Fung, D. S. S., … Verhulst, F. C. (2011). International comparisons of behavioral and emotional problems in preschool children: Parents' reports from 24 societies. *Journal of Clinical Child and Adolescent Psychology*, 40(3), 456–467. https://doi.org/10.1080/15374416.2011.563472

Rockhill, C., Kodish, I., DiBattisto, C., Macias, M., Varley, C., Ryan, S. (2010). Anxiety disorders in children and adolescents. *Curr Probl Pediatr Adolesc* Health Care, 40, 66–99. https:doi.org/10.1016/j.cppeds.2010.02.002

Rolock, N., & Testa, M. F. (2005). Indicated child abuse and neglect reports: Is the investigation process racially biased. In D. Derezotes, J. Poertner & M. Testa (Eds.), Race matters in child welfare: The overrepresentation of African American children in the system (pp. 119-130). Washington, DC: Child Welfare League of America.

Romano, Weegar, K., Babchishin, L., & Saini, M. (2018). Cross-informant agreement on mental health outcomes in children with maltreatment histories: A systematic review. *Psychology of Violence*, *8*(1), 19–30. https://doi.org/10.1037/vio0000086

Rosanbalm, Snyder, E. H., Lawrence, C. N., Coleman, K., Frey, J. J., van den Ende, J. B., & Dodge, K. A. (2016). Child wellbeing assessment in child welfare: A review of four measures. *Children and Youth Services Review*, *68*, 1–16. https://doi.org/10.1016/j.childyouth.2016.06.023

Rosenfield, S. (2000). Gender and dimensions of the self: Implications for internalizing and externalizing behavior. In E. Frank (Ed.), *Gender and its effects on psychopathology* (pp. 23–36). American Psychiatric Publishing, Inc.

Sheldrick, R. C., Benneyan, J. C., Kiss, I. G., Briggs-Gowan, M. J., Copeland, W., & Carter, A. S. (2015). Thresholds and accuracy in screening tools for early detection of psychopathology. *Journal of Child Psychology and Psychiatry*, *56*(9), 936–948. https://doi.org/10.1111/jcpp.12442

Shin, N.U., & Hong, S. (2019). Comparing the robustness of stepwise mixture modeling with continuous nonnormal distal outcomes. *Educational and Psychological Measurement,* 79(6), 1156–1183. https://doi.org/10.1177/0013164419839770

Shore, N., Sim, K.E., Le Prohn, N.S., Keller, T.E. (2002). Foster parent and teacher assessments of youth in kinship and non-kinship foster care placements: Are behaviors perceived differently across settings? *Child Youth Services Review*, 24, 109–134. https://doi.org/10.1016/S0190-7409(01)00170-0

Smith, S. (2007). Making sense of multiple informants in child and adolescent psychopathology: A guide for clinicians. *Journal of Psychoeducational Assessment*, *25*(2), 139–149. https://doi.org/10.1177/0734282906296233

Strijker, J., van Oijen, S., & Knot-Dickscheit, J. (2011). Assessment of problem behaviour by foster parents and their foster children: Assessment of problem behaviour. *Child & Family Social Work,* 16(1), 93–100. *https://doi.org/10.1111/j.1365-2206.2010.00717.x*

Tarren-Sweeney, M. J., et al. (2004). Are foster parents reliable informants of children's behaviour problems? *Child: Care, Health & Development*, 30, 2, 167–75. https://doi.org/10.1111/j.1365-2214.2003.00407.x

Tanaka, M., Wekerle, C., Schmuck, M. L., & Paglia-Boak, A. (2011). The linkages among childhood maltreatment, adolescent mental health, and self-compassion in child welfare

adolescents. *Child Abuse & Neglect*, *35*(10), 887–898. https://doi.org/10.1016/j.chiabu.2011.07.003

Tang M.H., Pinsky, E.G. (2015). Mood and affect disorders. *Pediatric Review*, 36, 52–60. https://doi.org/10.1542/pir.36-2-52

Thakur, H., & Cohen, J. (2019). Depression screening in youth: Multi-informant algorithms for the child welfare setting. *Psychological Assessment*, *31*(8), 1028–1039. https://doi.org/10.1037/pas0000728

Thomas, R., Sanders, S., Doust, J., Beller, E., & Glasziou, P. (2015). Prevalence of attention-deficit/hyperactivity disorder: A systematic review and meta-analysis. *Pediatrics (Evanston),* 135(4), e994–e1001. https://doi.org/10.1542/peds.2014-3482

van de Schoot, R., Sijbrandij, M., Winter, S. D., Depaoli, S., & Vermunt, J. K. (2017). The GRoLTS-Checklist: Guidelines for reporting on latent trajectory studies. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(3), 451-467. https://doi.org/10.1080/10705511.2016.1247646

van Dulmen, M. H. M., & Egeland, B. (2011). Analyzing multiple informant data on child and adolescent behavior problems: Predictive validity and comparison of aggregation procedures. *International Journal of Behavioral Development*, *35*(1), 84–92. https://doi.org/10.1177/0165025410392112

Vasileva, M., & Petermann, F. (2018). Attachment, development, and mental health in abused and neglected preschool children in foster care: A meta-analysis. *Trauma, Violence & Abuse,* 19(4), 443–458. https://doi.org/10.1177/1524838016669503

Villodas, M. T., Litrownik, A. J., Newton, R. R., & Davis, I. P. (2016). Long-term placement trajectories of children who were maltreated and entered the child welfare system at an early age: Consequences for physical and behavioral well-being. *Journal of Pediatric Psychology*, *41*(1), 46–54. https://doi.org/10.1093/jpepsy/jsv031

Xu, Y., & Bright, C.L. (2018). Children's mental health and its predictors in kinship and non-kinship foster care: A systemic review. *Children and Youth Services Review, 89*, 243-262. https://doi.org/10.1016/j.childyouth.2018.05.001

Youngstrom, Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of Consulting and Clinical Psychology*, *68*(6), 1038–1050. https://doi.org/10.1037/0022-006X.68.6.1038

Woods, Farineau, H. M., & McWey, L. M. (2013). Physical health, mental health, and behaviour problems among early adolescents in foster care. *Child Care, Health & Development*, *39*(2), 220–227. https://doi.org/10.1111/j.1365-2214.2011.01357.

Wulczyn, F., Chen, L., & Orlebeke, B. (2008). Evaluating contract agency performance in achieving reunification. *Children & Youth Services Review,* 31, 506-512. https://doi.org/10.1016/j.childyouth.2008.10.006

VITA

Dr. Jennifer Osborne earned her doctoral degree in Clinical Psychology from Loyola University Chicago. Before attending Loyola University Chicago, she attended Elon University in Elon, North Carolina, where she earned a Bachelor of Arts on Sociology and Strategic Communications in 2016. She then worked as a contractor for the Children's Bureau in Washington D.C. curating resources for the Child Welfare Information Gateway and writing content for National Foster Care Month.

While at Loyola, Jennifer gained extensive research experience under the guidance of Dr. Scott Leon in the Promoting Adjustment in Children through Evaluation (PACE) Lab. As a member of the PACE lab, she explored the role of social support, interrater reliability, and related mental health outcomes among youth involved in the child welfare system. Her work has appeared in numerous publications, including the *Journal of Traumatic Stress,* the *Journal of Child and Family Studies*, and the *Journal of Public Child Welfare*. Jennifer also received the Research Mentoring Award and Frank Kobler Travel Award in recognition of her research during her time at Loyola University Chicago.

Currently, Dr. Osborne is a Clinical Psychology Resident at the Children's Hospital of Michigan specializing in Pediatric Neuropsychology. She will complete a post-doctoral fellowship pediatric neuropsychology at NorthShore University Health System and continue her work to exploring the relationship between social support and mental health among youth with complex medical and neurodevelopmental histories.