UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Discrepancies in Reporting of Behavior Problems in Children With Autism Spectrum Disorder: Contribution of Child, Parent and Teacher Characteristics

Permalink

https://escholarship.org/uc/item/7089r470

Author

Llanes, Elizabeth

Publication Date

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Discrepancies in Reporting of Behavior Problems in Children with Autism Spectrum
Disorder: Contribution of Child, Parent, and Teacher Characteristics

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Elizabeth Lara Llanes

June 2019

Dissertation Committee:
      Dr. Jan Blacher, Chairperson
      Dr. Katherine Stavropoulos
      Dr. Keith Widaman

The Dissertation of Elizabeth Lara Llanes is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

Dedication

This work is dedicated to my husband, Peter Barnes, who has supported me, encouraged me, and endured many sacrifices towards helping me achieve my goals. I am humbled by the love and care you have given me while I experienced the ups and downs of my graduate career. I could not have completed this project without the unending support and confidence you had in me.

Thank you to my advisor and mentor, Dr. Jan Blacher, for believing in me and encouraging me to reach my potential as a graduate student. You have always supported me and had my back throughout my years at SEARCH and throughout the dissertation process. Thank you as well to my other committee members, Dr. Katherine Stavropoulos and Dr. Keith Widaman for your ongoing support and feedback throughout the dissertation process.

Finally, thank you to my colleagues at the SEARCH Family Autism Resource Center for your friendship, mentorship, and support over the years. You were my sounding board and my cheerleaders throughout this process.

ABSTRACT OF THE DISSERTATION

Discrepancies in Reporting of Behavior Problems in Children with Autism Spectrum
Disorder: Contribution of Child, Parent, and Teacher Characteristics

by

Elizabeth Lara Llanes

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside, June 2019
Dr. Jan Blacher, Chairperson

Despite the high rates of both internalizing and externalizing behavior problems in

children with ASD (Simonoff et al., 2008; van der Meer et al., 2012), there are frequently

inconsistencies between informants on behavioral rating scales, particularly when rating

internalizing behavior (De Los Reyes et al., 2015; Stratis & Lecavalier, 2015).

Discrepancies between raters can hinder treatment, as children, parents, and teachers

frequently fail to agree on the target problem, making it difficult for all informants

participating in treatment to work together to mitigate the problem (De Los Reyes &

Kazdin, 2005). The aims of the current study are to examine discrepancies between

parent and teacher ratings of behavior over time in a sample of young children with ASD,

and to identify parent, child, and teacher characteristics associated with informant

discrepancies. Participants in this study included 180 children with ASD ages 4 to 8 years

old and their parents and teachers. Internalizing and externalizing behavior problems

were measured using parent and teacher report on the Child Behavior Checklist and the

corresponding Teacher Report Form (CBCL and TRF; Achenbach & Rescorla, 2001).

Latent growth modeling was used to examine developmental trajectories of parent and

teacher ratings of child behavior, as well as to examine informant discrepancies across

time. The results indicate that, on average, parents tend to identify more problem

behaviors, particularly internalizing behaviors. No significant changes in parent-teacher

discrepancies were found across time. The magnitudes of parent-teacher discrepancies in

behavior ratings as well as child, parent, and teacher characteristics which predict the

discrepancies were also examined. Implications for research and practice are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Discrepancies in reporting of behavior problems in children with autism spectrum

disorder: Contribution of child, parent, and teacher characteristics

Children with autism spectrum disorder (ASD) experience significantly more

behavior problems than do their typically developing (TD) peers (Baker & Blacher, 2015;

Bauminger, Solomon, & Rogers, 2010). In fact, studies have found that about 70% of

children with ASD experience at least one comorbid psychiatric disorder (Gillberg &

Fernell, 2014; Leyfer, 2006; Mattila et al., 2010; Simonoff et al., 2008). Behavior

problems are frequently categorized as internalizing behavior problems (i.e., anxiety or

mood-related problems) and externalizing behavior problems (i.e., disruptive behaviors).

Children with ASD tend to experience high levels of internalizing problems. In particular,

a meta-analysis conducted by van Steensel, Bogels, and Perrin (2011) found that across

studies, about 40% of youth with ASD had at least one comorbid anxiety disorder. The

prevalence rate of mood disorders in youth with ASD is variable in the literature (1.4% to

38%; Magnuson & Constantino, 2011). However, Kim et al. (2000) found that rates of

depression are higher for youth with ASD when compared to their TD peers. Rates of

externalizing behavior problems are also relatively high in youth with ASD. Baker and

Blacher (2015) found that youth with ASD experienced significantly higher levels of

disruptive behavior disorders when compared to TD peers, and met diagnostic criteria for

ADHD at four times the rate of TD youth. ADHD, one of the disorders most frequently

comorbid with ASD, has been found to occur in anywhere from 50-80% of youth with

ASD (Simonoff et al., 2008; van der Meer et al., 2012).

Despite the high rates of both internalizing and externalizing behavior problems in children with ASD, inconsistencies between informants, or lack of inter-rater agreement, on behavioral rating scales frequently occurs, particularly when rating internalizing behavior (Achenbach, 1987; De Los Reyes et al., 2015; Grietens et al., 2004; Kanne, Abbachi, & Constantino, 2009; Salbach-Andrae, Lenz, & Lehmkhul, 2009; Stranger & Lewis, 1993; Stratis & Lecavalier, 2015; Youngstrum, Loeber, & Stouthamer-Loeber, 2000). Discrepancies between raters can hinder treatment, as children, parents, and teachers frequently fail to agree on the target problem, making it difficult for all informants participating in treatment to work together to address the behavioral concern (De Los Reyes, 2013). Informant discrepancies also have a particular relevance to clinical research, because evidence supporting the efficacy of treatments for children depends on multi-informant reports of child outcomes (Weisz, Jensen Doss, & Hawley, 2005). This study will examine parent and teacher discrepancies in ratings of internalizing and externalizing behavior problems in children with ASD over multiple time points. It will also investigate child, parent, and teacher characteristics which may influence discrepancies in parent-teacher ratings. Parent-teacher ratings in particular will be the focus of the current study due to the young age of participants (ages 4 to 8) as well as the fact that youth with ASD struggle to provide reliable reports of their own behavior problems (Mazefsky, Kao, & Oswald, 2011).

Parent and Teacher Behavior Rating Discrepancies

   Typically Developing Children. Parent and teacher behavioral rating discrepancies are consistent across the literature. De Los Reyes and colleagues (2015)

conducted a meta-analysis of 341 studies published between 1989 and 2014 to examine cross-informant correspondence in ratings of child behavior. To be included in the study, the published articles had to focus on internalizing and externalizing behaviors in children under 18. De Los Reyes et al. found low-to-moderate agreement across informant pairs, consistent with the findings of the meta-analysis conducted by Achenbach (1987) years prior. According to the authors, ratings of externalizing behavior had greater correspondence (mean $r = .30$; ps $<.001$) than ratings of internalizing behavior (mean $r = .25$; ps $<.001$) across rater pairs. Multiple informants' reports were also found to share little variance with one another. De Los Reyes et al. also found different magnitudes of agreement for different pairs of raters. For instance, mothers and fathers showed greater correspondence than parent-child or parent-teacher rater pairs.

Numerous studies have examined parent and teacher agreement on the measure that will be used to assess internalizing and externalizing behavior problems in the current study, the Child Behavior Checklist – Parent and Teacher Report Forms (CBCL and TRF, respectively; Achenbach & Rescorla, 2001). Even when utilizing different methods for calculating informant agreement, research has demonstrated low-to-moderate agreement among informants and found that agreement is better for externalizing problems than internalizing problems (Grietens et al., 2004; Salbach-Andrae et al., 2009; Stranger & Lewis, 1993; Youngstrum et al., 2000). These findings held for studies that examined informant agreement of ratings of young children (ages 5-6; Grietens et al., 2004) as well as studies examining ratings of older children (ages 11-18; Salbach-Andrae et al., 2009; Stranger & Lewis, 1993; Youngstrum et al., 2000).

Several studies found that, when determining whether a child was experiencing clinically significant internalizing and externalizing problems, agreement was very low (Youngstrum et al., 2000; Grietens et al., 2004). In fact, based on conditional probabilities for agreement on whether a child had clinically significant internalizing or externalizing behavior problems, Stranger and Lewis (1993) found that no informant's ratings could substitute for those of another rater.

Children with ASD. Similar to the literature examining parent and teacher agreement in TD children, previous research has consistently identified lack of agreement between parent and teacher report of problem behaviors in children with ASD. Stratis and Lecavalier (2015) conducted a meta-analysis of 49 studies examining informant agreement for youth (under age 22) with ASD or intellectual disability (ID). They found that, for both groups (ASD and ID), parents and teachers demonstrated moderate levels of informant agreement (mean weighted effect size across raters and behaviors of 0.36) and that agreement was higher for externalizing problem behaviors ($r = .42$) than for internalizing behaviors ($r = .35$). Although there was no difference between youth with ASD and ID for externalizing behavior, agreement on internalizing behavior was significantly higher for youth with ASD than youth with ID across raters (ASD: $r = .32$; ID $r = .29$, $Q = 4.31$, $p = .04$).

Correlates of Parent/Teacher Rating Discrepancies: Child and Informant Characteristics

Typically Developing Children. In their review of informant discrepancies in the assessment of child disorders, De Los Reyes and Kazdin (2005) identified various child

and parent characteristics that were associated with informant discrepancies. Across studies, there was evidence that gender is related to lack of rater agreement, with informants rating girls more consistency than boys. Child age and race were also associated with discrepancies between raters. Specifically, informants tended to have higher agreement when rating younger children than adolescents (Achenbach et al., 1987; De Los Reyes & Kazdin, 2005; Schroeder et al., 2010). For parent-teacher ratings specifically, Van Dulmen and Egeland (2011) found that discrepancies between parent and teacher ratings of externalizing behavior were smaller for younger children (at age 7) than for adolescents (at age 16). This is because teacher ratings of problem behaviors tend to decrease over time, whereas parent reports tend to be more stable over time, leading to larger parent-teacher discrepancies. The same is true for internalizing behavior – rater discrepancies tend to grow larger as the child matures into adolescence (van der Ende, Verhulst, & Tiemeier, 2012). When looking at child race, De Los Reyes and Kazdin found greater agreement for Caucasian children compared to African American children.

De Los Reyes and Kazdin (2005) also found that parent characteristics influenced informant discrepancies, including socioeconomic status (SES) and parent mental health. Although SES has long been a well-established predictor of children's behavior (Dodge, Pettit, & Bates, 1994; Bradley & Corwyn, 2002), the association of SES with informant discrepancies was mixed (De Los Reyes & Kazdin, 2005). Youngstrom et al. (2000) found no relationship between SES, based on maternal employment and education, and parent-teacher rating discrepancies. However, other researchers have found that lower

family income was related to larger discrepancies between parent and teacher ratings of child behavior (Collishaw et al., 2009; Stone et al., 2013). Zahner and Daskalakis (1998) also found that agreement between mothers and teachers was higher if mothers had higher levels of education. With regards to parent mental health, De Los Reyes and Kazdin found that maternal depression played a role in parent-teacher agreement, with depressed mothers being more likely to over-report child behavior problems (De Los Reyes et al., 2008 Richters, 1992; Youngstrom et al., 2000). This relation is consistent with the depression-distortion hypothesis, which posits that depression leads the informant to view the child's behavior with a negative bias (Richters & Pellegrini, 1989). Mother stress has also been related to parent-teacher rating discrepancies (Youngstrom et al., 2000), with greater maternal stress leading to larger differences between raters.

Only two studies have investigated the relations between teacher characteristics and multi-informant ratings of child behavior in typically developing children. Zahner and Daskalakis (1998) examined behavior problems in a population-based sample of over 1,400 children, ages six to eleven years, as rated by their parents and teachers, using the Child Behavior Checklist (CBCL) and Teacher Report Form (TRF), respectively (Achenbach & Rescorla, 2001). They found that parents tended to give higher ratings than teachers, especially on the internalizing scale, and that agreement was low (Internalizing: $\kappa = 0.14$; Externalizing: $\kappa = 0.33$). Teacher familiarity and contact with the child were significant predictors of parent-teacher agreement for both the internalizing and externalizing scales, indicating that teachers rate more consistently with parents when they know the child better.

Berg-Nielsen and colleagues (2012) also examined teacher characteristics as determinants of the parent-teacher agreement on ratings of behavior problems in preschool children. Using a sample of over 700 Norwegian children and their parents and teachers, Berg-Nielsen et al. (2012) found similar results to those of Zahner and Daskalakis (1998). Teachers provided lower ratings of children's behavior problems than did parents, especially on internalizing problems. Conflict in the student-teacher relationship, as measured by the Student-Teacher Relationship Scale (Pianta, 2001), was also a significant predictor of disagreement between parent and teacher ratings of internalizing behavior problems, with greater conflict predicting higher discrepancies. Teachers who reported higher levels of conflict in the student-teacher relationship tended to identify more problem behaviors, and although Berg-Nielsen and colleagues found that parents tended to rate more problem behaviors than teachers in general, conflict in the student-teacher relationship actually led teacher ratings to surpass parent ratings and predicted even higher informant discrepancies. In the Berg-Nielsen et al. (2012) study, teachers also tended to rate girls' behavior problems much lower than boys' behavior problems. These findings led Berg-Nielsen and colleagues to conclude that there were possible teacher perception biases at play in rating behavior problems, particularly when there is conflict in the student-teacher relationship, when identifying internalizing problems, and in the presence of gender preference for girls versus boys.

Children with ASD. The literature has also identified several moderators specific to informant discrepancies in the rating of behavior problems in children with ASD. For instance, in their meta-analysis, Stratis and Lecavalier (2015) found that age and IQ were

significant moderators of informant discrepancies between informants. For internalizing problems, parents and teachers showed higher agreement when rating adolescents than school-aged children. When rating social skills, agreement between parents and teachers was higher for school-aged children than for adolescents. Stratis and Lecavalier also found that IQ was a significant moderator of informant discrepancies. With increasing IQ, agreement among all raters on internalizing problem behaviors and total problem behaviors decreased. Stratis and Lecavalier believe this could indicate that youth with lower IQ have less variability in their behavior across environments, leading to increased agreement between raters.

The type of behavior problem being assessed also impacts the magnitude of informant discrepancies. Compared to ratings of TD youth, Kanne et al. (2009) found that parents and teachers were most discrepant when rating internalizing behavior, particularly anxiety-related symptoms.

Other studies have examined parent characteristics which may impact informant discrepancies in behavioral ratings of children with ASD. For instance, parent age and education level were associated with higher ratings on measures of child ASD symptoms, indicating that parents who were more educated were more likely to endorse problem behaviors when rating child behavior (Hattier, Matson, Belva, & Adams, 2013); perhaps they were more informed about specific behaviors associated with ASD. Reed and Osborne (2013) explored the correlation between parent stress and informant agreement. Although they found that stressed parents tended to produce slightly higher ratings of behavior problems in their children, stress had little impact on parent-teacher behavior

rating discrepancies, indicating a need to identify teacher characteristics that may impact informant agreement (Reed & Osborne, 2013). However, no work currently found in the literature describes the examination of teacher characteristics that correlate with informant discrepancies in ratings of children with ASD.

Parent-Teacher Rating Discrepancies as Predictors

Rater disagreement can often impact the identification of problems and treatment; therefore, a growing number of studies have examined the predictive power of rater discrepancies on a variety of child outcomes. In his review, De Los Reyes (2011) found eleven studies that examined informant discrepancies as predictors of outcomes such as delinquency, treatment response, and parent involvement in treatment. However, only one of these studies examined the impact of parent-teacher discrepancies on outcomes related to child psychopathology. Ferdinand, Van der Ende, and Verhulst (2007) examined parent-teacher ratings of behavior problems on the CBCL and TRF in 1,154 four to twelve-year-old children and then measured child outcomes 14 years later. Whereas scores on the CBCL and TRF predicted many later life mental health and behavioral outcomes, parent-teacher rating discrepancies for the most part did not. There were a few exceptions; for example, parent-teacher discrepancies in ratings of aggressive behaviors in the children were associated with an increased risk of suicide attempts and/or self-harm in adulthood. In addition, parent-teacher discrepancies when rating anxious/depressed symptoms in the children predicted child mood disorders at fourteen-year follow-up. Ferdinand et al. (2007) believe that parent-teacher rating discrepancies may be indicative of lack of home-school support or communication, which results in

under-identification of children's problems. Left untreated, these problems can lead to detrimental outcomes later in life.

Why are Parent and Teacher Ratings So Discrepant?

The literature on informant discrepancies has made it evident that differences in rater reports are due to much more than measurement error (Achenbach, 2001; De Los Reyes, 2011; De Los Reyes et al., 2015). Two factors that may contribute to parent and teacher rating discrepancies are rater-bias and context (De Los Reyes & Kazdin, 2005). Rater bias may be due to the fact that raters might have different motivations for rating, or different perspectives on what is typical versus abnormal behavior. However, we know that children's behavior can vary across settings and this may indicate that parents and teachers are both providing valuable information that can be used to inform intervention (De Los Reyes, 2011).

To compare parent and teacher reports of disruptive behaviors to clinical observations, De Los Reyes and colleagues (2009) developed an experiment where children were brought into a clinical setting and observed while interacting with a parent or an unfamiliar clinical examiner. They found that their observations of the child's disruptive behavior were consistent with variations in reports from their parents and teachers. For example, when a child was observed to be disruptive with a parent but not with the clinical examiner, it was more likely that parents reported high levels of disruptive behaviors while teachers reported low levels. Similarly, if a child was disruptive with the clinical examiner but not with the parent, this corresponded to higher ratings of disruptive behaviors from teachers than from parents. When both parents and

teachers reported disruptive behaviors, this was seen in the laboratory observations in child interactions with the parent and clinical examiner. De Los Reyes et al. (2009) concluded that variations in child's disruptive behaviors reflected contextual factors - interaction with a parent versus a non-parental adult.

Consistent with the findings of De Los Reyes and colleagues (2009), Hartley et al. (2011) found that parent-teacher discrepancies in reports of aggressive behavior were related to rater's perceptions of the environmental cues that triggered the aggressive behavior (e.g., task demands placed on the child or negative interaction with a peer). In fact, they found that the more similar the environments within which informants rated aggressive behavior, the more similar the reports from informants. This finding is consistent with behavioral theory, which posits that children are more likely to exhibit a behavior if certain contingencies (or environmental cues) are present.

To explore further the role of contextual factors, De Los Reyes et al. (2013) conducted the first experimental study to examine whether informants (mothers and children) would be able to provide context-specific information on child and family behavior. Families were randomly assigned to a receive a program that trains informants to use setting information when reporting (Setting-Sensitive Assessment) or to a control condition. The Setting-Sensitive Assessment training resulted in greater differences between mother-reported and child-reported behavior than the control group, suggesting that discrepancies arose because informants incorporated unique, context-specific information when rating behavior.

11

Some research has been conducted to tease apart the root of parent-teacher report discrepancies in ASD samples. Kanne et al. (2009) found disparities between parent and teacher reports of problem behaviors in children with ASD on the CBCL and TRF in a sample of 177 children ages 3 to 18 years (mean age of 7.3 years). At first, they hypothesized that parents and teachers may be observing the same behaviors, but rating them at different levels of severity; however, their analysis indicated that this was not the case. To determine this, Kanne and colleagues used bivariate correlations for each of the CBCL symptom domains and found low correlations, especially for internalizing behavior, suggesting that parent-teacher lack of agreement was not simply due to informant ratings of the same problems at differing levels of severity. Kanne et al. were also able to rule out rater bias by comparing the inter-rater agreement between parents and teachers when rating the behaviors of youth with ASD versus their typically developing (TD) siblings. Inter-rater agreement was much higher when parents and teachers were rating the TD siblings, suggesting that rater bias did not account for the discrepancy in the rating of youth with ASD. Kanne and colleagues suggested that these findings provided evidence that, for youth with ASD, problem behaviors were manifested differently across environmental contexts. Applied to the current study, these findings suggest that informant discrepancies were a result of the observation of different behaviors at home and at school. It might also be that typically developing children are easier to rate than children with ASD, leading to discrepancies between parent and teacher ratings for children on the autism spectrum.

Theoretical Models of Informant Discrepancies

Several theoretical models have been posited to attempt to provide explanations and methods for understanding discrepancies in informant ratings. The situational specificity (SS) perspective holds that children behave differently in different environments, as noted above. Therefore, this model attributes informant discrepancies to the contextual behavior of the child being rated. There are numerous studies which have found evidence for the contextual specificity of behavior, giving support to this theoretical perspective (Biglan, 1995; Morris, 1988). However, this theory does not take into account rater biases which might impact informant agreement.

To address rater biases, the multitrait-multimethod (MTMM) perspective was developed (Campbell & Fiske, 1959). This theory views different raters as different methods of measurement. Discrepancies between raters reflect systematic biases, or distorted reports of the behavior. For instance, biases such the halo effect, where one positive impression of a person leads the rater to perceive the person's other behaviors or qualities positively, impacts informant ratings. Viewing informant discrepancies through the lens of the MTMM model allows researchers to qualify the measurement error due to rater biases. Therefore, in the MTMM model, parent-teacher rating discrepancies give us more information about the characteristics of the rater than about the child's behavior across settings. The other issue with the MTMM perspective is that it does not account for informant discrepancies caused by different informants' knowledge of a child's behavior in different environments.

Building on the MTMM perspective, Kraemer and colleagues (2003) developed a theory to explain rater discrepancies. They proposed four dimensions that result in variability between raters: (1) the actual behavior being measured (T), (2) the context in which the behavior is observed (C), (3) the informant's perspective or characteristics that influence ratings (P), and (4) measurement error (E). Each rating is the sum of these four dimensions: T + C + P + E. Based on this theory, when informant perspectives are highly similar, or the behavior being rated is very consistent across settings, or the raters are observing behavior in the same context, raters will be less discrepant. In order to reduce rater discrepancies, Kraemer and colleagues recommend that researchers and clinicians focus on reducing extraneous variance due to different rater perspectives and contextual factors. To reduce variation that is due to rater perspective, Kramer et al. recommend additional informants in each context (e.g., mother and father ratings of the behavior at home; ratings of teacher and classroom aide in the school). To reduce extraneous variance due to contextual factors, Kramer et al. recommend gaining ratings from the same perspective (same rater) across settings. For instance, this would mean the parent observes and rates the child's behavior at school and, vice versa, the teacher rates the child's behavior at home. These recommendations might well reduce unexplained variation in their model, but one must call into question the feasibility of obtaining multiple rater reports across multiple settings. Another weakness of this theory is that it assumes that rater discrepancies are undesirable. However, other theories take a different perspective, namely, that differences between raters are meaningful and provide clinically

14

useful information about the child's behavior across settings. One such theory was developed by De Los Reyes and Kazdin (2005).

Building upon the SS and MTMM perspectives, De Los Reyes and Kazdin (2005) developed the Attribution-Bias-Context (ABC) model, which provides a framework for conceptualizing informant discrepancies across rater pairs (e.g., mother-teacher, mother-child). The researches identified three sources of informant discrepancies: the actor-observer phenomenon, perspective and memory recall, and the context in which the behavior occurs. The actor-observer phenomenon is the principle that one tends to attribute one's own behavior to environmental context while attributing the other's behavior to dispositional qualities (Jones & Nisbett, 1972). Therefore, informant discrepancies arise, in part, due to the fact that informants disagree in their attribution of the causes of the child's behavior problems. Based on the actor-observer phenomenon, De Los Reyes and Kazdin hypothesized that parents and teachers would be more likely to attribute the child's behavior to dispositional factors (within child factors), while children would more likely attribute their behavior to the context in which the behavior occurred (environmental factors) when providing self-ratings of their behavior.

Another source of informant discrepancy is informant perspective and memory recall. Because observer informants (parents and teachers) are more likely to attribute behavior problems to within child factors, they are more likely than the child to recall negative behaviors that are consistent with their perspective (Tversky & Marsh, 2000). In contrast, children are more likely to attribute their behavior to environmental factors and therefore recall the contextual factors that led to the behavior, consistent with their

perspective. Therefore, the ABC model posits that ratings from observer informants (parents and teachers) will be more similar when compared to each other than when compared to child ratings.

Finally, the third component ABC model is the context in which the child's behavior is observed. Informant discrepancies may arise when parents and teachers perceive different behaviors in different contexts. For instance, a parent could observe oppositional behavior in the home as the primary problem, whereas the child's teacher observes inattention at school as the problem. But De Los Reyes and Kazdin (2005) emphasized that, although it is likely that parents and teachers observe different behaviors in different settings, attention must be paid to informants' perceptions and their recall of the child's behavior from memory in any theoretical conceptualization of informant discrepancies.

Another proposed theory to bridge the SS and MTMM perspectives was developed by Dumenci, Achenbach, and Windle (2011). Using hierarchical constructs, they proposed a hybrid perspective that takes into account the environmental context and the perspective of the informant. Their measurement design includes three facets: item (specific behaviors), syndrome (e.g., internalizing problems, externalizing problems), and informant (e.g., mother, teacher). These facets are used to estimate contextual factors (i.e., behaviors specific to the context from the perspective of an informant) and cross-contextual components (i.e., behaviors that are common across contexts and informants).

Dumenci et al. (2011) also identified three systematic sources of variability in ratings, including (1) a cross-contextual higher-order trait common across all three facets

of measurement, (2) a contextual higher-order trait common across the item and syndrome facets but different for each informant, and (3) a contextual lower-order trait specific to all possible combinations of syndrome and informant facets. All other sources of variability not represented in these three sources are considered random error of measurement. To illustrate their model, they used the example of parent and teacher ratings of child externalizing behavior. The first source of variance, the cross-contextual higher-order trait, would then refer to the externalizing behaviors manifested across contexts (home and school) as reported by multiple informants (parents and teachers). The second source of variation would be due to the fact that the child also exhibits externalizing behaviors in a particular setting as reported by one informant. The third source of variation in the model, the contextual lower-order trait, refers to a specific type of externalizing behavior (e.g., aggression) manifested in a specific setting (e.g., home) as reported by a specific informant (e.g., mother).

The model proposed by Dumenci et al. (2011) has many strengths, including the fact that these researchers developed a parametric representation of the model which they tested with data from mother, teacher, and child ratings of rule-breaking and aggressive behavior. However, their model does not separate the contextual aspects of behavior from rater bias. Rather, when they use the term "contextual," Dumenci and colleagues are referring to setting-specific behavior from the perspective of the informant.

Validity of Informant Discrepancies

When examining the validity of multi-informant approaches in their review, De Los Reyes et al. (2015) focused on incremental validity, or whether the addition of a new

informant would increase the predictive power beyond the data supplied by just one informant. They noted that the best evidence in support of the incremental validity of a multi-informant approach is for externalizing behavior problems because assessments that evaluate these behaviors can be corroborated with direct observation. Few studies have examined the validity of multi-informant reports for internalizing behaviors, such as anxiety (De Los Reyes et al., 2015). This is, in part, due to the complexities of operationalizing and observing internalizing behaviors. Because internalizing behaviors are more difficult to observe, we also expect reliability to be lower across raters (De Los Reyes et al., 2013).

In addition to the evidence that indicates how data from multiple raters adds incremental validity in the assessment of externalizing behavior, there is also emerging evidence that indicates that discrepancies among informants can predict treatment responses. For instance, De Los Reyes, Alfano, and Beidel (2010) administered a pre-treatment parent and self-report measure that assessed social phobia symptoms in 81 children (ages 7 to 16) and then provided a treatment for social phobia. They found a significant relation between treatment response and parent-child disagreement, with larger informant discrepancies associated with non-response to treatment. This study provides evidence that disagreement between raters can not only inform variability in treatment outcomes, but also inform treatment response.

Use and Interpretation of Informant Discrepancies. The clinical use and interpretation of informant discrepancies are also essential to any discussion concerning their validity. Multi-informant ratings are used either to make diagnostic decisions or to

inform treatment. Information from multiple reporters can be treated a number of ways. One approach is the conjunctive or converging approach, which requires multiple informants to agree on the presence of a behavior or syndrome (Youngstrum et al., 2003; De Los Reyes et al., 2013). For instance, if determining whether or not the child is experiencing depressive symptoms, a clinician using the conjunctive approach would require the parent, child self-report, and teacher reports of depressive symptoms to be in the clinical range for depression to be considered present in the child. This strategy can overlook children with complex symptom presentations or children who exhibit problem behaviors only in a particular setting. Understandably, the conjunctive approach results in lower identification rates of children with behavior problems, and therefore is considered less sensitive. De Los Reyes et al. (2013) noted that this approach is frequently used in research, as opposed to practice, which is problematic given that informants' reports frequently give diverging information and that these disagreements yield important information regarding the behavior being assessed.

Another approach is the disjunctive or diverging approach, in which only one informant must endorse a behavior problem to meet the diagnostic threshold (Youngstrum et al., 2003; De Los Reyes et al., 2013). This approach is the most commonly used in clinical practice and results in the broadest pool of identified children. Although this approach yields a high sensitivity, in that it correctly identifies a large percentage of individuals who are presenting with a behavior problem, it also has low specificity in that it yields a high false-positive rate. It has been argued that this approach to synthesizing multi-informant data is valid due to the high reliability of behavior

checklists as well as their cross-validation with other rating scales (Achenbach, 1995). The reliability and validity of these scales can be interpreted as evidence that informants are providing reliable and valid reports of child behavior across different settings.

The third and final approach described by Youngstrum and colleagues (2003) is the compensatory approach. In this approach, scores from multiple raters are summed or averaged to create one combined score. Although Youngstrum et al. hypothesized that this method would be the most psychometrically sound given that an aggregate measure would reduce individual error and possess greater reliability, they found the opposite. They collected parent, teacher, and self-report ratings and DSM-IV categorizations of nearly 200 youth being treated for mood or disruptive behavior disorders. The compensatory approach (the average score across raters) resulted in lower agreement with the DSM-IV classification than the disjunctive approach, which compared agreement between individual raters and the DSM-IV categorization.

Operational Triad Model. De Los Reyes, Thomas, Goodman, and Kundey (2013) made an attempt to bridge the current approaches to dealing with informant discrepancies in research through their Operations Triad Model (OTM). The OTM allows researchers to test whether meaningful differences in the behavior assessed can be identified when interpreting discrepancies among multiple raters. To assess multi-informant data, most clinical research uses Converging Operations, or a set of measurement conditions within which one assesses the validity of multi-informant reports based on their similarity (or convergence) with one another. For instance, an intervention study looking at behavioral outcomes in children might claim that there is more evidence of behavior change if seen

by more than one rater. However, the OTM creates a framework from which researchers can view discrepancies as meaningful and informative, rather than as a set of findings that weaken the conclusions that may be drawn.

The OTM first requires researchers to pose a priori hypotheses regarding whether informants are expected to agree or disagree. If the expectation is disagreement, then a hypothesis is also made to determine what the diverging findings will reflect (e.g., variation in the behavior across settings or measurement error). Then, based on the reports of the informants, the pattern of convergence or divergence is interpreted. This model is dependent upon the current research literature which overwhelmingly indicates that informant discrepancies are largely due to valid representations of variation in children's behavior across settings (De Los Reyes et al., 2009; De Los Reyes et al., 2013; Hartley et al., 2011).

If a researcher expects that the informants' reports will yield converging information, he or she can test whether or not Converging Operations account for the study's findings. If informant reports are diverging, then the researcher tests for Diverging Operations, or whether or not the conflicting reports are indicative of a meaningful difference in the behavior being assessed. Specifically, the researcher tests whether or not each report meets the a priori reliability threshold set by the researcher or if the informant ratings relate to other constructs that would support the validity of their reports. The researcher should also rule out methodological factors in the study that might explain informant discrepancies. If measurement error or methodological problems account for the informant discrepancies, then this is indicative of Compensating

Operations. Using post-hoc tests, the researcher can test for Compensating Operations and can use statistical approaches to correct for differences between reports and examine convergence among informants' reports. If the issue was due to invalid or unreliable reports from a subset of the sample, the researcher can also exclude these reports from the sample in favor of reports that yield reliable and valid data.

According to De Los Reyes et al. (2013), the OTM is valuable because it involves making hypotheses about the presence of informant discrepancies in research conclusions. It encourages researchers to use multi-informant data, even when such data might result in conflicting results, because it provides a framework for attributing discrepancies to meaningful differences in the behavior assessed. In the future, De Los Reyes and colleagues believe that the use of OTM to interpret commonly discrepant reports of parent and teacher reports of child externalizing behavior might result in the development of standardized batteries of parent, teacher, and independent observer measures to tease apart cross-contextual versus context-specific expressions of child externalizing behaviors.

Measurement of Informant Discrepancies

The way in which researchers operationally define, or measure, informant discrepancies is crucial and can drastically effect the magnitude of the discrepancies depending on the methods used.

Measuring Informant Agreement. Instead of conceptualizing differences in informant ratings as discrepancies, some researchers have measured informant agreement. The most prevalent method for examining rater agreement is Pearson's

product moment coefficient, *r*, a bivariate correlation measuring the linear relationship between reports from two informants (e.g., Youngstrom et al., 2000). Another approach is to use the intraclass correlation coefficient (ICC) as a measure of agreement. According to Cichetti (1994), this method is less dependent on systematic effects for one rater over another (e.g., mothers systematically rating children higher than teachers), and therefore is considered a better measure of agreement.

Another approach to operationalizing informant agreement is to use *Q* correlations. *Q* correlations are Pearson *r* correlations between sets of common items provided by different raters (Achenbach, 2011). Achenbach (2011) notes that *Q* correlations can inflate agreement, even if one rater consistently rates items higher than the other rater, because it is not affected by differences between the raters' metrics. However, this also means that *Q* correlations are useful for calculating agreement between raters who see children in environments with different levels of the problem behavior (e.g., at home and school).

Measuring Discrepancies Between Informants. In addition to methods for evaluating informant agreement, other statistics have been developed to measure differences between raters. In their review, De Los Reyes and Kazdin (2004) compared three common ways of measuring discrepancies between informants: raw difference scores, standardized difference scores, and residual difference scores. When using raw difference scores, one rater's unstandardized, raw score is subtracted from the other rater's raw score. In the standardized difference score, raters' scores are first converted into z-scores based on the distribution of informant ratings in the sample. The advantage

of the standardized difference score is that it places informant ratings on the same metric and makes the resulting scores more comparable and interpretable. Finally, in using residual difference score, one informant's rating is used as the independent variable in a regression model and used to predict the other informant's rating. The residual difference is the standardized difference between the predicted value and the informant's actual rating.

De Los Reyes and Kazdin (2004) compared these three measurement methods using a sample of mothers, who each rated her child's externalizing behavior. Using the three different difference score methods yielded different conclusions, depending on the variation of informant ratings and the correlation between informants. For instance, if the variation in one group of informants is much larger than the variation in another group of raters, then raw difference scores and standardized difference scores produce widely different findings particularly when relating the difference scores to other informant characteristics (such as SES, maternal psychopathology, child age, etc.). De Los Reyes and Kazdin concluded that future studies should use the standardized difference score because that was the only difference score measurement method that was correlated with each of the informants' ratings.

Some studies have called the use of difference scores into question. Laird and De Los Reyes (2013) noted that using difference scores makes the interpretation of discrepancies challenging, even when the values are squared (i.e., using directional difference score) or absolute (i.e., taking absolute value of a difference score). They also add that difference scores do not adequately determine whether informant discrepancies

can predict psychopathology or whether they can be predicted from other constructs, such as maternal depressive symptoms.

If the research question of interest is the relation between informant discrepancies and an outcome variable of interest, Laird and Weems (2011) recommended using a regression model, regressing the outcome variable on the separate scores of two informants (P = parent scores; T = teacher scores):

$$Y = \beta_0 + \beta_1 P + \beta_2 T + e$$

where β_{1r} is the parameter estimate for the association between parent's report of the behavior and the outcome when holding teacher's report constant, and β_{2r} is the parameter estimate for the association between teacher's report of the behavior and the outcome, holding parent's report constant. Although this model is useful when addressing research questions examining the impact of informant discrepancies on an outcome, it is not helpful if the outcome of interest is the actual informant discrepancy.

Research examining informant discrepancies as outcomes have primarily used difference scores (De Los Reyes & Kazdin, 2004) and have not examined discrepancies longitudinally. Whereas Laird and De Los Reyes (2013) cautioned against the use of difference scores, other work by Rogosa and Willett (1983) demonstrated that difference scores are reliable in most situations, particularly when reliability estimates of the original scores are good and when there is variability in individuals' difference scores. However, it is also important to note that most research examining difference scores have used observed scores. Difference scores using observed measures provide valuable information but they are still subject to measurement error, which is compounded given

the fact that the observed values also carry error. This led to the assertion that "the difference between two fallible measures is frequently much more fallible than either" (Lord, 1963, p. 32). The advantage to using a structural equation modeling (SEM) framework is that it utilizes latent variables, which are free of measurement error. Latent differences can then be estimated using latent true score values (free of error), which are estimated from the observed scores. The resulting latent difference estimates are more reliable and less prone to measurement error than difference scores derived from observed values (McArdle, 2009). Given these advantages, this study will explore SEM techniques to model similarities and differences between latent scores representing parent and teacher true scores on child behavior rating scales. This will provide information regarding the magnitude and direction of informant discrepancies.

Research Questions

Children exhibiting behavior problems are commonly referred for psychological assessment. As a part of the assessment process, clinicians frequently turn to behavior rating scales, such as the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), which have strong validity and reliability support in the literature. However, best practice in behavioral and psychological assessment involves the use of multi-informant data (Merrell, 2007); problems arise when data gathered from multiple sources are contradictory. Discrepant data can make it challenging to interpret the child's presenting problems and can sometimes delay or even limit access to treatment. The research literature examining informant discrepancies in ratings of child behavior has attempted to guide clinicians and researchers alike in how to interpret disparate rater reports validly

(De Los Reyes et al., 2015). Nonetheless, there is still much work to be done when considering special populations, such as children with ASD. Thus, this study will examine informant discrepancy among parents and teachers in a sample of young children with ASD.

There is some evidence that, among youth with ASD, informant agreement is moderated by child age (Stratis & Lecavalier, 2015); however, no study was found that examined informant discrepancies longitudinally in a sample of young children with ASD. In addition, although some child and parent characteristics have been found to correlate with informant agreement, teacher characteristics have not been examined in ratings of children with ASD (Reed & Osborne, 2013). The current study attempts to address these gaps in the literature and poses the following research questions:

1. Do parent reports of child internalizing and externalizing behavior problems change over time and across age?

2. Do teacher reports of child internalizing and externalizing behavior problems change over time and across age?

3. How discrepant are parent and teacher reports of of child internalizing and externalizing behavior problems? Are parent-teacher rating discrepancies consistent over time?

4. Which child (age, gender, IQ), parent (income), and teacher (years teaching experience, class size, role as general or special education teacher, autism experience, and student teacher relationship) characteristics predict parent and

teacher discrepancies in reports of child internalizing and externalizing behavior problems?

Methods

Participants

Participants in this study will include 184 children ages 4 to 7 years old (150 boys, 34 girls), and their parents and teachers, who were participating in a longitudinal study across two sites in Southern California and Massachusetts. The purpose of the larger study was to examine factors that lead to a successful school transition for young children with ASD. Participants were referred to the study by local service agencies, schools, and state regional centers. Three cohorts of students were recruited, and all participations were subject to the same recruitment, eligibility, and study procedures.

To meet enrollment criteria for the larger study, participants were required to have a previous diagnosis of an autism spectrum disorder. Those who did not have a previous diagnosis or who only had a school designation of autistic-like were also administered the Autism Diagnostic Interview – Reivsed (ADI-R; LeCouteur, Lord, & Rutter, 2003). In addition, participants were required to have an intelligence quotient (IQ) on the Wechsler Preschool and Primary Scales of Intelligence (WPPSI-III; Wechsler, 2002) of 50 or above. All children were screened as part of the study with the Autism Diagnostic Observation Schedule for Children (ADOS-2, Lord et al., 2008). Children were classified under autism or autism spectrum if they met ADOS score cutoff criteria. Mean participant scores on the WPPSI-III were within average levels of cognitive functioning

on the FSIQ ($M = 87.71$, $SD = 17.73$), with 17.8% having comorbid intellectual disability (IQ at or below 70).

The sample was predominantly male (81.7%) and White (62.8%). Only 40.3% of parents reported household incomes under $65,000. Mother education in this sample was defined by years of completed schooling. Overall, most mothers in our sample completed high school (97.2%), with more than half of mothers reporting completion of a college degree or higher (63.9%). The majority of child participants were enrolled in public elementary schools (59.2%) and public preschool programs (10.1%), with only a small percentage (3.3%) enrolled in private schools. *Table 1* contains participant demographic information.

Procedure

The Institutional Review Boards of the participating universities approved study procedures. Informed consent forms were mailed home, reviewed with parents and collected on the day of the child's first assessment, after reviewing the form again. In nearly all cases (88.3%), the participating parent was mother. The parents completed measures of social skills and child behavior problems prior to each visit. Once deemed eligible, children were assessed during the fall (Time 1) and spring (Time 2) of the same school year, and the winter (Time 3) of the following school year. During the visits, parents were also asked to give consent for their child's teacher to provide information about the child and his or her school environment; they also completed a demographic and other forms or questionnaires. Once consent and contact information were obtained,

teachers were mailed measures to complete. *Table 2* describes participant sample size (parents and teachers) across the time points.

During the on-site assessment, graduate student researchers trained in study procedures met separately with the child and mother to complete a variety of tasks. Activities with the mother included an interview on topics related to the child's behavior, relationships with his or her teacher and peers, school experiences, and overall transition to school. Assessments of child behavior problems were obtained via mother- and teacher-completed questionnaires.

Measures

Background information from the parent and child was obtained via a parent completed demographic questionnaire completed at the eligibility visit. Teachers also completed a short Classroom Climate Inventory questionnaire about their teaching experiences and demographics. Behavioral information was collected from parents and teachers at Time 1, Time 2 and Time 3. The following measures were given to all children in this study.

Eligibility: Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 2008). The ADOS is a standardized, semi-structured play-based observation of child behavior in situations that elicit autistic tendencies. Four modules can be administered, dependent on the child's verbal ability. The observation yields scores in four domains: Social Interaction, Communication, Stereotyped Behaviors and Restricted Interests, and Play. Of these domains, only two, Social Interaction and Communication, are included in the algorithm for determining the child's overall ADOS

score. Individuals can fall into one of three categories resulting from this score: autism, autism spectrum, or not on the autism spectrum. These categories are predetermined cutoff points provided in the ADOS manual based on the specific module that is being administered. The ADOS has established reliability and validity from research on a sample of children with a diagnosis of autism (Lord et al., 2008). The ADOS has high discriminative validity with high sensitivity (97%, 95%, and 90% across Modules 1 to 3, respectively) and specificity (94%, 87%, and 94%, across Modules 1 to 3, respectively) in discriminating between children with ASD and children without a spectrum disorder.

Elligibility: Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III; Wechsler, 2002). Children's cognitive skills were measured with the WPPSI-III. The WPPSI-III is composed of 14 subtests and yields an IQ score scaled as a standard score ($M = 100$, $SD = 15$). For this study, a calculated Full Scale IQ (FSIQ) score was computed from an abbreviated measure of cognitive functioning, which included three subtests: block design, matrix reasoning and vocabulary subscales. This instrument is intended for use with children between the ages of 2:6 and 7:3 years of age. The selection of these three subtests was based on their established reliability ($r = .95$) and high predictive validity in gaining an estimate of cognitive ability (Sattler & Dumont, 2004).

Achenbach System of Empirically Based Assessment (ASEBA) – Parent and Teacher Reports (CBCL and TRF; Achenbach & Rescorla, 2001). Behavior problems were measured using parent and teacher reports on the Achenbach System of Empirically Based Assessment (ASEBA) – Parent Report and the Teacher Report Form (CBCL and

TRF; Achenbach & Rescorla, 2001). Parents and teachers were asked to complete the items describing specific child behaviors on a three point Likert scale (0 = not true, 1 = somewhat or sometimes true, 2 = very true or often true). Higher scores on subscales indicate greater levels of problematic behaviors. Depending on the child's age, parents were either administered the CBCL for ages 1.5 to 5 or for ages 6 to 18. Teachers were administered the TRF for 1.5 to 5 or for ages 6 to 18. For the purposes of this study, only the Internalizing problem behavior scale and the Externalizing problem behavior scale will be used. A t-score of 70 or above indicates that the child is showing clinically elevated levels of behavior problems. The ASEBA scales were normed using a nationally representative sample of children and their parents and teachers. Achenbach and Rescorla (2001) demonstrated that the Internalizing and Externalizing problem scales discriminated between referred and non-referred samples and demonstrated convergent validity when compared to other behavioral rating scales, such as the Behavior Asseessment system for Children (BASC; Reynolds & Kamphaus, 1992) Scales (r's = .74 to .89). In this study, the Internalizing and Externalizing problem scales had good internal consistency (CBCL: Internalizing α's = .82-.94, Externalizing α's = .88-.93; TRF: Internalizing α's = .83-.89, Externalizing α's = .92-.94).

Student Teacher Relationship Scale (STRS). The STRS (Pianta, 2001) is a 28-item questionnaire that measures teachers' perceptions of their relationship with specific students from preschool to grade 3. It yields a total score and three factors that measure Conflict (e.g., "This child easily becomes angry at me."), Closeness (e.g., "I share an affectionate, warm relationship with this child."), and Dependency (e.g., "This child asks

for my help when he/she really does not need help."). Responses were given on a 5-point

Likert scale from 1 ("definitely does not apply") to 5 ("definitely applies"). The STRS

has been shown to be correlated with current and future academic achievement, behavior

problems, and peer relationships (Birch & Ladd, 1998; Hamre & Pianta, 2001). In this

study, the STRS had good internal consistency for the total score (α's = .79-.82).

Data Analysis Plan

Descriptive statistics. The following analyses were computed using SPSS

Version 24.0 (IMB Corp, 2010). The data were first examined for outliers, which were

defined as values that fell three standard deviations above or below the mean. Outliers

were constrained to three standard deviations from the mean to reduce the influence of

extreme data points (Cohen et al., 2003). Descriptive statistics for parent- and teacher-

reported internalizing and externalizing behavior at each time point as well as

demographic information are provided in Table 1.

Research Question 1 and 2. The first two research questions ask whether parent

and teacher reports of child internalizing and externalizing behavior change over time and

across age. Latent growth curve models (LGMs) were used to examine the trajectories of

parent and teacher reports of child internalizing and externalizing behavior.

Across time: LGMs examine longitudinal data to address questions related to

developmental change across time. The basic LGM consists of two latent factors

(intercept factor and slope factor), with repeated measures of the construct over time as

indicators. Through LGMs, researchers can model intra-individual change across

individual intercept and slope estimates (Byrne, 2012). In the current study, four models

were examined: (1) parent-reported internalizing behavior, (2) parent-reported externalizing behavior, (3) teacher-reported internalizing behavior, and (4) teacher-reported externalizing behavior. These analyses were conducted using Mplus Version 7 (Muthen & Murthen, 2012).

Across age. Using LGM's, individual change in parent and teacher reported internalizing and externalizing behavior were modeled across the three time points of measurement. But because the three time points span only approximately 18-months, we also wish to model behavior problems across participant age for each of the measurement points so that we can examine developmental trajectories of these behavior problems. Participants entered the study at varying ages (4-7 years) and aged about 18-months between the first and last measurement points. Therefore, age in months was calculated for each participant and included in the model.

Model Specification. Model estimates were interpreted for the intercept and slope factors. The latent intercept factor is the starting value for the average participant, or, in other words, the predicted score at Time1. Factor loadings of the repeated measures are set to 1.0, which represents the starting point of the growth curve. The slope latent factor defines the slope of the growth curve and represents the rate of change of the trajectory over time. Therefore, the factor loadings for the slope estimates will be fixed to reflect each time of measurement. The three time points in the current study were not equally spaced; therefore, factor loadings were specified to match the time points of measurement throughout the 18-month period: Time 1 = 0, Time 2 = .5, and Time 3 = 1.5 (see Figures 1-4). For the LGMs in which behavior was modeled across age, age in months wastreated

continuously. Participants began the study between ages 48-92 months. Based on Mehta and West's (2000) recommendations for latent growth modeling with individually varying time points, participant age at Time 1 was added to the model as a predictor of the intercept and slope factors (see Figures 5-8). This accounted for the differences in the intercept and slope factors that are due to differences in age of participants at the first time point of measurement.

Variances of the intercept and slope latent factors reflect the variation of individuals around the overall group growth parameters. Because these models estimate the variability in individual change over time, they are random coefficients models.

Model Estimation. Multiple fit indices were examined to assess model fit. First, the Chi-square test statistic (χ^2) was examined as a global fit index, where non-significant values are indicative of good model fit. However, because this statistic is highly sensitive to sample size, alternative fit indices were also used. The root mean square error of approximation (RMSEA) is a measure of absolute fit based on the model degrees of freedom. Values between .05 and .08 indicate a good fit. Incremental indices, such as the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) were also examined to gauge comparative fit. Values above .90 are considered acceptable fit values (Byrne, 2012).

Research Questions 3 and 4. The third research question asks about the magnitude of the discrepancies between parent and teacher reports of child internalizing and externalizing behavior and whether or not the discrepancies are consistent over time. The fourth research question asks which child, parent, and teacher characteristics predict

parent and teacher discrepancies in reports of behavior across time. These questions were addressed using Structural Equation Modeling (SEM). SEM allows researchers to explore relationships between observed and unobserved (or latent) variables. Using SEM techniques, a latent factor representing similarities and one representing differences between parent and teacher reports of internalizing and externalizing behavior were specified using observed parent-teacher report values at each time point (see Figures 9 and 10). Then, using latent growth curve modeling, the similarity and difference factors were examined over time (see Figures 11-14). This allowed us to examine changes to parent-teacher discrepancies in behavior ratings across multiple time points. Covariates were added to the model to determine if child, parent, and teacher characteristics predict estimates of the difference factor.

Model Specification. Model estimates were interpreted for the similarities and difference factors based on observations of parent and teacher reports of internalizing and externalizing behavior at each time point. The similarities factor represents the magnitude to which parent and teacher reports of child behavior are alike. Factor loadings for the similarities factor are set to 1.0. The difference factor represents the magnitude to which parent and teacher reports are discrepant. Factor loadings for the difference factor were set to 1.0 for parent reports and -1.0 for teacher reports, indicating that teacher scores will be subtracted from parent scores. Figures 9 and 10 provide a visual representation of the models described.

Model estimates were also interpreted for the intercept and slope factors. For the both the similarity and difference models, the latent intercept factor is the starting

estimate for parent-teacher rating similarity for the average participant. Factor loadings of

the repeated measures are set to 1.0, which represents the starting point of the growth

curve. The slope latent factor represents the rate of change of parent-teacher rating

similarities or differences over time. Therefore, the factor loadings for the slope estimates

were fixed to reflect each time of measurement: Time 1 = 0, Time 2 = .5, and Time 3 =

1.5 (see Figures 11-14). Variances of the intercept and slope latent factors reflect the

variation among individuals around the overall group growth parameters.

Covariates. To determine which child, parent, and teacher characteristics predict

parent-teacher discrepancies in reports of behavior problems (research question 4),

covariates were added to the model to determine the impact of these characteristics at

each time point and across time. Child characteristics included age, gender, and IQ on the

WPPSI. Parent characteristics included level of parent education and racial and/or ethnic

background. Teacher characteristics included years of teaching experience, role as a

general or special education teacher, autism training experience, and student-teacher

relationship (STRS). Child and parent characteristics were measured only once (i.e., child

age, child gender, IQ, parent education, and parent race/ethnicity). However, all teacher

characteristics, except STRS, were measured twice, at Time 1 and Time 3, because Times

1 and 2 occurred during the same school year, whereas the third time point was during

the following school year. STRS was measured at all three time points. Therefore, teacher

characteristics were obtained from two different teachers. The covariates that did not vary

across time (i.e., child and parent characteristics) were modeled as predictors of the latent

slope and intercept factors for the difference factor over time. Teacher characteristics that

were taken at Time 1 were modeled as predictors of the difference factors at Time and Time 2 (same teacher). Teacher characteristics taken at Time 3 were modeled as predictors for the the difference factor at Time 3 (see Figures 15 and 16). Because STRS was taken at each time point, two parallel LGMs were created to model the slope and intercept factors for STRS and the difference factors across the three time points. Then the relations among the intercept and slope factors for the STRS LGM and the intercept and slope factors for the difference LGM were examined (see Figures 17 and 18).

Model Estimation. Like the other LGMs described previously, multiple fit indices were examined to assess the SEM and LGM model fits. The Chi-square test statistic (χ^2) was examined as a global fit index, where non-significant values are indicative of good model fit. The root mean square error of approximation (RMSEA), a measure of absolute fit based on the model degrees of freedom, was also used. Finally, incremental indices, such as the Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) were examined to gauge comparative fit.

Assumptions. Longitudinal modeling assumes that variables are measured across three or more time points, that measurement at each time occurred at the same interval across participants, and that scores have the same unit of measure and measure the same construct at each time point (Byrne, 2012). The data used for the current study met these assumptions. Participants were assessed at three time points. Although these three time points occurred with different spacing, we have specified model parameters to reflect the time between assessment points across an 18-month period.

To ensure that scores have the same unit of measure across time points, we needed to find a common metric for the internalizing and externalizing problem scales across forms. The CBCL and TRF are divided into two forms depending on child age (1.5-5 years and 6-18 years) and differ in terms of the number of items in the scale. Nevertheless, the forms measure the same constructs. The items are different due to the fact that internalizing and externalizing behaviors are expressed differently across age groups. The forms were created to reflect developmental changes in internalizing and externalizing problem behaviors between 1.5-5-year-old children and children 6 and over. Thus, the CBCL and TRF provide standardized scores, T-scores, for the internalizing and externalizing scales that are derived from large scale, population-based normative samples. Other studies have used T-scores to examine the ASEBA scales longitudinally (Green, Berkovits, & Baker, 2015).

In addition, we also used percentage of maximum possible (POMP) scores in all analyses, in addition to T-scores, and the results will bewere compared. POMP scores have many advantages; they are not sample or population dependent, and they allow comparisons across different measures of the same construct (Cohen, Cohen, Aiken, & West, 1999). They also produce scores that are very interpretable and easy to communicate to others.

Missing Data. The data were evaluated to determine if missing cases were related to the variables of interest in this study, namely behavior problems. If data are missing at random (MAR), this means that missing values depend on observed values and do not depend on unobserved variables (Little & Rubin, 2014). In this case, full maximum

likelihood (FIML) estimation will be used. Instead of using listwise or pairwise deletion, the FIML approach involves estimating a likelihood function for each individual based on using all the available data (Byrne, 2012). The advantage of FIML is that it has been shown to provide unbiased parameter estimates and standard errors while still retaining maximum statistical power that would be threatened if cases were deleted or data were treated as missing.

Results

Research Question 1: Parent Reports of Internalizing and Externalizing Behaviors

The first research question was related to changes in parent-reported child internalizing and externalizing behavior problems over time and across age. Latent growth modeling was used to investigate this question.

Internalizing Problems. Parent-reported internalizing problems in youth over time were explored using a latent growth model (see Figure 1). Because the three time points in the current study were not equally spaced, factor loadings were specified to match the time points of measurement throughout the 18-month period: Time 1 = 0, Time 2 = .5, and Time 3 = 1.5 (see Figures 1-4). The analysis of the latent growth curve models assessed mean intercept, mean slope, variance of intercept, variance of slope, and covariance between intercept and slope. To examine parent-reported internalizing problems across child age, age was added as a covariate to the model based on the recommendations of Mehta and West (2000).

T-Score LGM. The latent growth model using T-scores as the unit of measurement demonstrated good model fit, χ^2 (3) = 0.72, p = 0.87; RMSEA = .00, 90%

CI [.00-.06]; CFI = 1.00, TLI = 1.01. Based on the results displayed in *Table 3*, mean

parent-reported internalizing behavior at Time 1 was significantly different from zero at a

T-score of 63.32. The variance of the intercept also indicated that there was significant

variability in the initial rating by parents at time 1. The slope factor also indicated a

significant decrease in parent-reported internalizing behavior across time. However, there

was no significant variability in participants' change over time.

Next, child age in months was added to the model as a covariate. The addition of

this covariate did not impact model fit, $\chi^2(4) = 1.18$, $p = 0.88$; RMSEA = .00, 90% CI

[.00-.05]; CFI = 1.00, TLI = 1.02. Results of the latent growth model with age added as a

covariate are displayed in *Table 5*. Child age had a significant impact on the intercept

factor, indicating that the older the child was at Time 1, the lower their T-score on the

parent-reported internalizing problems scale. In contrast, child age had a significant

positive impact on slope, indicating that as the children in the study aged, parent-ratings

on the internalizing scale increased. This finding was surprising given that the latent

growth model examining parent-reported internalizing behavior across time demonstrated

a decrease over time.

POMP Score LGM. The POMP score latent growth model revealed similar

results to the T-score model. This model had good model fit, $\chi^2(3) = 5.95$, $p = 0.11$;

RMSEA = .07, 90% CI [.00-.16]; CFI = .99, TLI = .99. Results of the POMP score LGM

are displayed in *Table 3*. At time 1, parents reported a mean of 25% of the total possible

points on the CBCL internalizing problems scale. There was again significant variability

in the initial starting point for parent-reported internalizing problems. This model also

revealed a significant decrease in points over time and no significant variability in participants' change over time.

Child age in months was then added to the model as a covariate and the results are displayed in *Table 5*. After the addition of this covariate, model fit was still good, $\chi^2(4)$ = 6.17, p = 0.19; RMSEA = .06, 90% CI [.00-.13]; CFI = .99, TLI = .99. Like in the T-score model, child age had a significant negative impact on the intercept factor, indicating that the older the child was at Time 1, the lower their POMP score on the parent-reported internalizing problems scale. Just as in the T-score model, child age had a significant positive impact on slope, indicating that as the children in the study aged, parent-ratings on the internalizing scale slightly increased.

Overall, the T-score and POMP score models demonstrated very similar findings. Parent-reported internalizing behavior tended to increase across the 18-month period in which data was gathered but tended to decrease across child age. However, because the time span was only 18 months, the finding with regards to internalizing behavior across child age was considered to have more practical significance since child age at Time 1 spanned from 4 to 8 years old, a much larger window into child development.

Externalizing Problems. Parent-reported externalizing problems in youth over time were also explored using a latent growth model (see *Figure 2*). Factor loadings were again specified to match the time points of measurement throughout the 18-month period: Time 1 = 0, Time 2 = .5, and Time 3 = 1.5.

T-Score LGM. The latent growth model using T-scores as the unit of measurement demonstrated good model fit, $\chi^2(3)$ = 2.85, p = 0.42; RMSEA = .00, 90%

CI [.00-.12]; CFI = 1.00, TLI = 1.00. Based on the results displayed in *Table 3*, mean parent-reported externalizing behavior at Time 1 was significantly different from zero at a T-score of 60.39. The variance of the intercept also indicated that there was significant variability in the initial rating by parents at time 1. The slope factor also indicated a significant decrease in parent-reported externalizing behavior across time. There was significant variability in participants' change over time. There was also significant covariance between intercept and slope, -6.03, SE = 2.62, *p* = .02, indicating that those participants who started out with higher T-scores had smaller decreases over time than those who started out with lower T-scores.

Next, child age in months was added as a covariate (see *Table 5*). After the addition of this covariate, model fit was still good, $\chi^2(4) = 2.63$, *p* = 0.62; RMSEA = .00, 90% CI [.00-.09]; CFI = 1.00, TLI = 1.01. Child age was not a significant covariate for either the intercept factor or slope factor, indicating that child age in months did not impact T-scores on the parent-reported externalizing problems scale.

POMP Score LGM. The POMP score latent growth model revealed similar results to the T-score model. This model had good model fit, $\chi^2(3) = 2.85$, *p* = 0.42; RMSEA = .00, 90% CI [.00-.12]; CFI = 1.00, TLI = 1.00). Results of the POMP score LGM are displayed in *Table 3*. At time 1, parents reported a mean of 35% of the total possible points on the CBCL externalizing problems scale. There was again significant variability in the initial starting point for parent-reported externalizing problems. This model also revealed a significant decrease in points over time as well as significant variability in participants' change over time. Like in the T-score model, there was

significant covariance between the intercept and slope factors, -.004, SE = .001, p = .001,

indicating that those participants who started out with higher POMP scores had smaller

decreases over time than those who started out with lower POMP scores.

Child age in months was then added to the model as a covariate and the results are

displayed in *Table 5*. After the addition of this covariate, model fit was still good, $\chi^2(4)$

= 2.63, p = 0.62; RMSEA = .00, 90% CI [.00-.09]; CFI = 1.00, TLI = 1.01. In this model,

child age had a significant negative impact on the intercept factor, indicating that the

older the child was at Time 1, the lower their POMP score on the parent-reported

externalizing problems scale. Child age also had a significant positive impact on slope,

indicating that as the children in the study aged, parent-ratings on the externalizing scale

increased.

Overall, the T-score and POMP score models demonstrated similar findings.

Parent-reported externalizing behavior tended to increase across the 18-month period in

which data was gathered, however, for the POMP score model, parent ratings of

externalizing problems tended to decrease across child age. The T-score model

demonstrated no change in parent-reported externalizing behavior across child age.

Research Question 2: Teacher Reports of Internalizing and Externalizing Behaviors

Internalizing Problems. Teacher-reported internalizing problems were also

examined via latent growth modeling (see *Figure 3*). Factor loadings were specified to

match the time points of measurement throughout the 18-month period: Time 1 = 0, Time

2 = .5, and Time 3 = 1.5. Mean intercept, mean slope, variance of intercept, variance of

slope, and covariance between intercept and slope were also examined for teacher-reported problems.

   T-Score LGM. The latent growth model using T-scores as the unit of measurement demonstrated adequate model fit, $\chi^2(3) = 8.17$, $p = 0.04$; RMSEA = .10, 90% CI [.02-.19]; CFI = .93, TLI = .93. Based on the results displayed in *Table 4*, mean teacher-reported internalizing behavior at Time 1 was significantly different from zero at a t-score of 57.80. The variance of the intercept also indicates that there was significant variability in the initial rating by teachers at Time 1. The slope factor also indicated a significant decrease in teacher-reported internalizing behavior across time. There was significant variability in participants' change over time as well as significant covariance between the intercept and slope factors, indicating that those who started out with higher T-score ratings by teachers decreased less over time than those who were rated lower at Time 1.

   Next, child age in months was added to the model as a covariate. The new model also had adequate model fit, $\chi^2(4) = 8.08$, $p = 0.08$; RMSEA = .08, 90% CI [.00-.16]; CFI = .95, TLI = .92. Results of the latent growth model with age added as a covariate are displayed in *Table 5*. Child age was not a significant covariate for either the intercept factor or slope factor, indicating that child age in months did not impact T-scores on the teacher-reported internalizing problems scale.

   POMP Score LGM. Again, the POMP score latent growth model revealed similar results to the T-score model. This model also adequate model fit, $\chi^2(3) = 8.48$, $p = .04$; RMSEA = .10, 90% CI [.02-.19]; CFI = .92, TLI = .92. Results of the POMP score LGM

are displayed in *Table 4.* At Time 1, teachers reported a mean of 17% of the total possible

points on the CBCL internalizing problems scale. There was again significant variability

in the initial starting point for teacher-reported internalizing problems. This model also

revealed a significant decrease over time and significant variability in participants'

change over time. Like in the T-score model, there was significant covariation between

the intercept and slope factors in the same direction.

When child age in months was added as a covariate, model fit remained adequate,

$\chi^2(4) = 8.27$, $p = .08$; RMSEA = .08, 90% CI [.00-.16]; CFI = .95, TLI = .92. Results of

this model are displayed in *Table 5*. Child age had a significant negative impact on the

intercept factor, indicating that the older the child was at Time 1, the lower their POMP

score on the teacher-reported internalizing problems scale. Child age had no impact on

slope of teacher-reported POMP scores on the internalizing problems scale.

Overall, the T-score and POMP score models demonstrated very similar findings.

Teacher-reported internalizing behavior tended to increase across the 18-month period in

which data was gathered but tended to demonstrate no change across child age. However,

because the time span of the study was only 18 months, the finding that teacher-reported

internalizing behavior was static across child age was considered to have more practical

significance since child age at Time 1 spanned from 4 to 8 years old, a much larger

window into child development.

Externalizing Problems. Finally, latent growth modeling was used to examine

teacher-reported externalizing problems (see Figure 4). Factor loadings were again

specified to match the time points of measurement throughout the 18-month period: Time

1 = 0, Time 2 = .5, and Time 3 = 1.5.

T-Score LGM. The latent growth model using T-scores as the unit of

measurement demonstrated adequate model fit, $\chi^2(3) = 6.36$, $p = 0.04$; RMSEA = .11,

90% CI [.02-.22]; CFI = .96, TLI = .95. Based on the results displayed in *Table 4*, mean

teacher-reported externalizing behavior at Time 1 was significantly different from zero at

a T-score of 58.13. The variance of the intercept also indicates that there was significant

variability in the initial rating by teachers at Time 1. The slope factor also indicated a

significant decrease in teacher-reported internalizing behavior across time. There was

significant variability in participants' change over time as well as significant covariance

between the intercept and slope factors, indicating that those who started out with higher

T-score ratings by teachers decreased less over time than those who were rated lower at

Time 1.

Child age in months was then added to the model as a covariate. The addition of

this covariate resulted in poorer model fit, $\chi^2(4) = 18.02$, $p = 0.001$; RMSEA = .144,

90% CI [.08-.21]; CFI = .89, TLI = .83. Results of the latent growth model with age

added as a covariate are displayed in *Table 5*. Child age was not a significant covariate

for either the intercept factor or slope factor, indicating that child age in months did not

impact T-scores on the teacher-reported externalizing problems scale.

POMP Score LGM. Again, the POMP score latent growth model revealed similar

results to the T-score model. This model also had adequate model fit to the data, $\chi^2(3) =$

6.83, $p = .08$; RMSEA = .09, 90% CI [.00-.18]; CFI = .97, TLI = .97. Results of the

POMP score LGM are displayed in *Table 4*. At time 1, teachers reported a mean of 23% of the total possible points on the CBCL externalizing problems scale. There was again significant variability in the initial starting point for teacher-reported externalizing problems. Like in the T-score model, the POMP score model also revealed a significant decrease over time and significant variability in participants' change over time. Similar to the T-score model, there was significant covariation between the intercept and slope factors in the same direction.

When child age in months was added as a covariate, the model still had good fit statistics, $\chi^2(4) = 8.45$, $p = .08$; RMSEA = .08, 90% CI [.00-.16]; CFI = .97, TLI = .95. Results of this model are displayed in *Table 5*. Child age had significant negative impact on the intercept factor, indicating that the older the child was at Time 1, the lower their POMP score on the teacher-reported externalizing problems scale. Child age had no impact on slope of teacher-reported POMP scores on the externalizing problems scale.

Again, the T-score and POMP score models demonstrated very similar findings. Teacher-reported externalizing behavior tended to increase across the 18-month period in which data was gathered but tended to demonstrate no change across child age.

Research Question 3: Parent-Teacher Rating Discrepancies Over Time

The third research question investigated the magnitude of the discrepancies between parent and teacher reports of child internalizing and externalizing behavior and whether or not the discrepancies are consistent over time. This question was analyzed using structural equation modeling and latent growth curve modeling. Model estimates were interpreted for the similarities and difference factors based on observations of parent

48

and teacher reports of internalizing and externalizing at each time point. Factor loadings for the similarities factor are set to 1.0. Factor loadings for the difference factor were set to 1.0 for parent reports and -1.0 for teacher reports (see Figures 9 and 10). Then, latent growth curve models were used to examine the similarity and difference factors longitudinally. Model estimates were interpreted for the intercept and slope factors.

Internalizing Problems. Parent and teacher reports of internalizing behavior were examined (see Figures 11 and 12). Internalizing behavior ratings using T-scores were modeled first and then compared to the model using POMP scores.

T-Score Model. The latent growth curve model examining the similarity and difference factors over time had good model fit, $\chi^2(12) = 14.11$, $p = .29$; RMSEA = .03, 90% CI [.00-.09]; CFI = .99, TLI = .99). The mean of the difference factor at Time 1, the initial starting point of the LGM, was 2.80 T-score points. This indicates that, on average, parents' ratings on the internalizing behavior problems scale were 2.80 T-score points greater than teacher ratings. The standard deviation of the difference factor was 6.67 T-score points. The T-score latent growth curve model examining the difference factor across time revealed that there was no significant change over time in parent-teacher rating discrepancies on the internalizing problem behavior scale (see *Table 7*).

POMP Score Model. Using POMP scores, the latent growth curve model examining the similarity and difference factors over time also had good model fit, $\chi^2(11) = 16.36$, $p = .13$; RMSEA = .05, 90% CI [.00-.10]; CFI = .99, TLI = .98). The mean of the difference factor at Time 1, the initial starting point of the LGM, was .04 POMP score units. This indicates that, on average, parent ratings on the internalizing behavior

problems scale were 4% of the maximum possible points greater than teacher ratings. The

standard deviation of the difference factor was .09 POMP score units, or 9% of the

maximum possible points. Consistent with the T-score model, the POMP score latent

growth curve model examining the difference factor across time revealed that there was

no significant change over time in parent-teacher rating discrepancies on the internalizing

problem behavior scale (see *Table 7*).

Externalizing Problems. Parent and teacher reports of externalizing behavior

were examined next (see Figures 13 and 14). Externalizing behavior ratings using T-

scores were modeled first and then compared to the model using POMP scores.

T-Score Model. The latent growth curve model examining the similarity and

difference factors over time had adequate model fit, $\chi^2(11) = 22.95$, $p = .02$; RMSEA =

.08, 90% CI [.03-.12]; CFI = .97, TLI = .96. The mean of the difference factor at Time 1,

the initial starting point of the LGM, was 1.31 T-score points. This indicates that, on

average, parents' ratings on the externalizing behavior problems scale were 1.31 T-score

points greater than teacher ratings. The standard deviation of the difference factor was

5.61 T-score points. The T-score latent growth curve model examining the difference

factor across time revealed that there was no significant change over time in parent-

teacher rating discrepancies on the externalizing problem behavior scale (see *Table 7*).

POMP Score Model. The latent growth curve model examining the similarity and

difference factors over time had good model fit, $\chi^2(11) = 12.97$, $p = .30$; RMSEA = .03,

90% CI [.00-.09]; CFI = 1.00, TLI = .99. The mean of the difference factor at Time 1, the

initial starting point of the LGM, was .06 POMP score units. This indicates that, on

average, parent ratings on the externalizing behavior problems scale were 6% of the maximum possible points greater than teacher ratings. The standard deviation of the difference factor was .10 POMP score units, or 10% of the maximum possible points. Consistent with the T-score model, the POMP score latent growth curve model examining the difference factor across time revealed that there was no significant change over time in parent-teacher rating discrepancies on the externalizing problem behavior scale (see *Table 7*).

Research Question 4: Covariate Model

Next, child, parent, and teacher covariates were added to the model in a step-wise fashion to determine the impact of these characteristics on parent-teacher rating discrepancies.

Internalizing Problems. First internalizing problems were examined and the results of the T-score and POMP score models were compared. Child covariates were added first and non-significant relations were removed from the model before adding parent and then teacher covariates. This ensured that the final model was the most parsimonious.

T-Score Model. First, covariates of parent-teacher rating discrepancies in child internalizing behavior, measured using T-scores, were examined.

Child Characteristics. Child characteristics were examined first for their impact on the difference factor, which measures parent-teacher discrepancies in ratings of child internalizing behavior. The child characteristics examined included IQ, age, and gender. After inputting these covariates, the model continued to have good fit, $\chi^2(18) = 19.31, p$

51

= .37; RMSEA = .02, 90% CI [.00-.07]; CFI = 1.00, TLI = .99. Results of the model are displayed in *Table 8*. The only significant predictor of the intercept and slope of the difference factor was child IQ, Intercept on IQ: 0.24, SE = .06, z = 4.00, p < .001; Slope on IQ: -.10, SE = .09, z = -2.1, p = .03. This indicates that higher child IQ predicted greater discrepancies between parent and teacher reports of internalizing behavior at Time 1. However, higher IQ also predicted less change over time in the difference factor across time points. Child age and gender were not significant predictors of the intercept or slope of the difference factor and were removed from the model. When the model was rerun without child age and gender, child IQ ceased to be a predictor of the slope of the difference factor, Slope with IQ: -.04, SE = .05, z = -.95, p = .34, and this relation was subsequently removed from the model as well.

 Parent Characteristics. Next, parent income was added to the model. After including this parent characteristic, the model retained good fit to the data, $\chi^2(23)$ = 27.38, p = .24; RMSEA = .03, 90% CI [.00-.07]; CFI = .99, TLI = .98. Results of the model are displayed in *Table 8*. Parent income was only a predictor of intercept of the difference factor, -.79, SE = .37, z = -2.15, p = .03, indicating that higher parent income predicted lower parent-teacher rating discrepancies in child internalizing behavior at Time 1. Parent income was not a significant predictor of the slope of the difference factor and this pathway was subsequently removed from the model.

 Teacher Characteristics. Next, teacher characteristics (i.e., professional training in autism, years of teaching experience, role as a general education or special education teacher, and class size) were examined in the model. Because teacher characteristics were

only measured at Times 1 and 3, they did not vary across time. Therefore, they were modeled as predictors of the difference factor at Time 1 and 3, respectively. When added, the model fit statistics decreased but were still adequate, $\chi^2(58) = 79.65$, $p = .03$; RMSEA = .08, 90% CI [.02-.11]; CFI = .89, TLI = .84. The results of the model are displayed in *Table 9*. None of the teacher characteristics were predictors of the difference factor at Time 1 or Time 3.

   Final Model: Adding The Student Teacher Relationship. Because the STRS scale was measured at each time point, two parallel LGMs were created to model the slope and intercept factors for STRS and the difference factors across the three time points of the study. Then the relationship between the intercept and slope factors for the STRS LGM were examined. The results of the STRS LGM are displayed in *Table 10*. This STRS model, on its own, had good model fit, $\chi^2(3) = 4.62$, $p = .21$; RMSEA = .06, 90% CI [.00-.15]; CFI = .97, TLI = .97. When the STRS LGM was added to the model, the larger model maintained good model fit, $\chi^2(48) = 69.97$, $p = .02$; RMSEA = .05, 90% CI [.02-.08]; CFI = .95, TLI = .94. The relationship between the latent STRS intercept and latent STRS slope on the latent difference factor intercept and slope are displayed in *Table 11*. The latent STRS intercept had a positive relationship with the latent difference factor intercept, 45.78, SE = 8.40, $z = 5.45$, $p < .001$, and a negative relationship with the latent difference factor slope, -22.74, SE = 6.00, z = -3.79, $p < .001$. This indicates that, at Time 1, the stronger the relationship between students and teachers, the more discrepant parents and teachers were in their ratings of internalizing problem behaviors. However, across time, stronger student-teacher relationships were related to decreased parent-

teacher discrepancies in ratings of internalizing behavior. The latent STRS slope had a negative relationship with the latent difference factor intercept, -17.19, SE = 6.71, z = -2.56, p = .01, and a positive relationship with the latent difference factor slope, 17.20, SE = 5.00, z = 3.44, p = .001. This indicates that increases in the student-teacher relationship across time were associated with lower initial parent-teacher discrepancies in ratings of internalizing behavior at Time 1. Increases in the student-teacher relationship across time were also related to increases in parent-teacher discrepancies in ratings of internalizing behavior across time.

Results of the final model are displayed in *Figure 17*. In this final step of the model, higher child IQ was still predictive of higher parent-teacher discrepancies in internalizing behavior at Time 1, .05, SE = .02, z = 2.27, p = .02. Higher parent income was still predictive of lower parent-teacher discrepancies in ratings of internalizing behavior at Time 1, -.26, SE = .12, z = -2.14, p = .03.

POMP Score Model. Next, a model using POMP scores was examined and compared to the results of the T-score model.

Child Characteristics. After adding child covariates (IQ, age in months, and gender) to the model, the model continued to have good fit, $\chi^2(17)$ = 20.96, p = .37; RMSEA = .04, 90% CI [.00-.08]; CFI = .99, TLI = .98. Results of the model are displayed in *Table 8*. Child IQ was a significant predictor of the intercept and slope of the difference factor, Intercept on IQ: 0.003, SE = .001, z = 3.88, p < .001; Slope on IQ: -.001, SE = .001, z = -2.02, p = .04. This indicates that higher child IQ predicted greater discrepancies between parent and teacher reports of internalizing behavior at Time 1.

Higher IQ also predicted less change over time in the difference factor across time points. Child gender was also a significant predictors of the slope of the difference factor, .05, SE = .02, z = 1.97, p < .05, indicating that being female was predictive of an increase in the difference factor slope, or in other words, an increase in parent-teacher discrepancies in ratings of internalizing across time. Child age was not a significant predictor of the intercept or slope of the difference factor.

  Parent Characteristics. Next, parent income was added to the model. After including this parent characteristic, the model maintained good fit statistics, $\chi^2(25) =$ 25.26, p = .45; RMSEA = .01, 90% CI = .00-.06; CFI = 1.00, TLI = 1.00. Results of the model are displayed in *Table 8*. Like in the T-score model, parent income was only a predictor of intercept of the difference factor, -.01, SE = .01, z = -1.98, p < .05, indicating that higher parent income predicted lower parent-teacher rating discrepancies in child internalizing behavior at Time 1. Parent income was not a predictor of the slope of the difference factor and this pathway was subsequently removed from the model.

  Teacher Characteristics. Next, teacher characteristics (i.e., professional training in autism, years of teaching experience, role as a general education or special education teacher, and class size) were examined as predictors of the difference factor at Time 1 and 3. When added, the model fit statistics decreased slightly but were still adequate, $\chi^2(59)$ = 69.70, p = .16; RMSEA = .05, 90% CI [.00-.10]; CFI = .95, TLI = .93. The results of the model are displayed in *Table 9*. Like in the T-Score model, in the POMP score model, none of the teacher characteristics were predictors of the difference factor at Time 1 or Time 3.

Final Model: Adding The Student Teacher Relationship. When the STRS LGM was added to the model, the larger model maintained good model fit, $\chi^2(52) = 66.05$, $p = .09$; RMSEA = .04, 90% CI [.00-.07]; CFI = .97, TLI = .96. The relationship between the latent STRS intercept and latent STRS slope on the latent difference factor intercept and slope are displayed in *Table 11*. Like in the T-score model, the latent STRS intercept had a positive relationship with the latent difference factor intercept, .51, SE = .11, $z = 4.81$, $p < .001$, and a negative relationship with the latent difference factor slope, -.25, SE = .07, $z = -3.57$, $p < .001$. This indicates that, at Time 1, the stronger the relationship between students and teachers, the more discrepant parents and teachers were in their ratings of internalizing problem behaviors. However, across time, stronger student-teacher relationships were related to decreased parent-teacher discrepancies in ratings of internalizing behavior. Similar to the T-score model, the latent STRS slope had a negative relationship with the latent different factor intercept in the POMP score model, -.18, SE = .09, $z = -2.05$, $p = .04$, and a positive relationship with the latent difference factor slope, .18, SE = .06, $z = 3.26$, $p = .001$. Just as with the T-score model, this indicates that in the POMP score model, increases in the student-teacher relationship across time were associated with lower initial parent-teacher discrepancies in ratings of internalizing behavior at Time 1. Increases in the student-teacher relationship across time were also related to increases in parent-teacher discrepancies in ratings of internalizing behavior across time.

Results of the final model are displayed in *Figure 18*. Like in the T-score final model, higher child IQ was still predictive of higher parent-teacher discrepancies in

internalizing behavior at Time 1, .001, SE = .00, z = 2.86, p = .004. Child IQ ceased to be

a predictor of the latent difference factor slope and so this relationship was dropped from

the model at this stage. Higher parent income was also predictive of lower parent-teacher

discrepancies in ratings of internalizing behavior at Time 1, -.004, SE = .002, z = -2.55, p

= .01. Child gender, specifically being female, was also predictive of an increase in the

difference factor slope, or in other words, an increase in parent-teacher discrepancies in

ratings of internalizing across time, .02, SE = .002, z = 2.56, p = .01.

Externalizing Problems. Externalizing problems were examined next and the

results of the T-score and POMP score models were compared. Like in the internalizing

problem models, child covariates were added first and insignificant relationships were

removed from the model before adding parent and then teacher covariates. This ensured

that the final model was the most parsimonious.

T-Score Model. First, covariates of parent-teacher rating discrepancies in child

externalizing behavior, measured using T-scores, were examined.

Child Characteristics. Child characteristics (IQ, age, and gender) were examined

first for their impact on the difference factor, which measures parent-teacher

discrepancies in ratings of child externalizing behavior. After inputting these covariates,

the model continued to have adequate fit, $\chi^2(17)$ = 32.66, p = .01; RMSEA = .07, 90%

CI [.03-.10]; CFI = .97, TLI = .94. Results of the model are displayed in *Table 8*. The

only significant predictor of the intercept of the difference factor was child IQ, 0.18, SE =

.05, z = 3.35, p = .001. This indicates that higher child IQ predicted greater discrepancies

between parent and teacher reports of externalizing behavior at Time 1. Child IQ did not

significantly predict the slope of the difference factor and this pathway was removed from the model. Similarly, child age and gender were not significant predictors of the intercept or slope of the difference factor and were also removed from the model.

Parent Characteristics. Next, parent income was added to the model. After including this parent characteristic, the model had good fit statistics, $\chi^2(23) = 33.73$, $p = .07$; RMSEA = .05, 90% CI [.00-.09]; CFI = .98, TLI = .97. Results of the model are displayed in *Table 8*. Parent income was a predictor of the intercept of the difference factor, -.62, SE = .32, $z = -1.95$, $p = .03$, indicating that higher parent income predicted lower parent-teacher rating discrepancies in child internalizing behavior at Time 1. Parent income was not a predictor of the slope of the difference factor and this pathway was subsequently removed from the model.

Teacher Characteristics. Then, teacher characteristics (i.e., professional training in autism, years of teaching experience, role as a general education or special education teacher, and class size) were examined in the model. These characteristics were modeled as predictors of the difference factor at Time 1 and 3. When added, the model maintained good fit statistics, $\chi^2(52) = 61.53$, $p = .17$; RMSEA = .05, 90% CI [.00-.10]; CFI = .95, TLI = .92. The results of the model are displayed in *Table 9*. None of the teacher characteristics were predictors of the difference factor at Time 1 or Time 3.

Final Model: Adding The Student Teacher Relationship. Finally, the relationship between the intercept and slope of the difference factor and the intercept and slope of the STRS LGM were examined. When the STRS LGM was added to the model, the model maintained adequate fit, $\chi^2(39) = 63.15$, $p = .01$; RMSEA = .06, 90% CI = .03-.09; CFI

= .96, TLI = .95). The relations of the latent STRS intercept and latent STRS slope with the latent difference factor intercept and slope are displayed in *Table 11*. The latent STRS intercept had a positive relationship with the latent difference factor intercept, 38.42, SE = 7.25, $z = 5.30$, $p < .001$, and a negative relationship with the latent difference factor slope, -17.07, SE = 5.23, $z = -3.27$, $p = .001$. This indicates that, at Time 1, the stronger the relationship between students and teachers, the more discrepant parents and teachers were in their ratings of externalizing problem behaviors. However, across time, stronger student-teacher relationships at Time 1 were related to a decrease in the trajectory of parent-teacher discrepancies in ratings of externalizing behavior. The latent STRS slope had a negative relation with the latent difference factor intercept, -20.31, SE = 5.98, $z = -3.40$, $p = .001$, and a positive relation with the latent difference factor slope, 21.12, SE = 4.60, $z = 4.59$, $p < .001$. This indicates that increases in the student-teacher relationship across time were associated with lower initial parent-teacher discrepancies in ratings of externalizing behavior at Time 1. Increases in the student-teacher relationship across time were also related to increases in the trajectory of parent-teacher discrepancies in ratings of externalizing behavior across time.

Results of the final model are displayed in *Figure 19*. In this final step of the model, higher child IQ was no longer a significant predictor of higher parent-teacher discrepancies in externalizing behavior at Time 1, .03, SE = .02, $z = 1.51$, $p = .13$. Higher parent income was still predictive of lower parent-teacher discrepancies in ratings of externalizing behavior at Time 1, -.30, SE = .16, $z = -2.42$, $p = .02$.

POMP Score Model. Next, a model using POMP scores was examined and compared to the results of the T-score model.

Child Characteristics. After adding child covariates (IQ, age in months, and gender) to the model, the model continued to have good fit, $\chi^2(17) = 22.39$, $p = .17$; RMSEA = .04, 90% CI [.00-.08]; CFI = .99, TLI = .98. Results of the model are displayed in *Table 8*. Child IQ was a significant predictor of the intercept the difference factor, 0.003, SE = .001, $z = 3.04$, $p = .002$. This indicates that higher child IQ predicted greater discrepancies between parent and teacher reports of externalizing behavior at Time 1. IQ was not a significant predictor of the slope of the difference factor for externalizing problems. However, child age in months was a significant predictor of the intercept and slope of the difference factor, Intercept on Age: -.006, SE = .001, $z = -4.07$, $p < .001$; Slope on Age: .002, SE = .001, $z = 2.18$, $p = .03$. This indicates that increasing age was predictive of a decrease in the difference factor intercept, or in other words, a decrease in parent-teacher discrepancies in ratings of externalizing behavior at Time 1. It also indicates that increasing age was also predictive of higher rates of growth in parent-teacher discrepancies in ratings of externalizing behavior across time. Child gender was not a significant predictor of the intercept or slope of the difference factor.

Parent Characteristics. Next, parent income was added to the model. After including this parent characteristic, the model maintained good fit statistics, $\chi^2(21) = 31.13$, $p = .07$; RMSEA = .05, 90% CI [.00-.09]; CFI = .98, TLI = .97). Results of the model are displayed in *Table 8*. Like in the T-score model, parent income was only a predictor of intercept of the difference factor, -.01, SE = .01, $z = -2.07$, $p = .04$, indicating

that higher parent income predicted lower parent-teacher rating discrepancies in child externalizing behavior at Time 1. Parent income was not a predictor of the slope of the difference factor and this pathway was subsequently removed from the model.

   Teacher Characteristics. Next, teacher characteristics (i.e., professional training in autism, years of teaching experience, role as a general education or special education teacher, and class size) were examined as predictors of the difference factor at Time 1 and 3. When added, the model fit statistics decreased slightly but were still adequate, $\chi^2(55) = 61.72$, $p = .25$; RMSEA = .04, 90% CI = .00-.09; CFI = .97, TLI = .95. The results of the model are displayed in *Table 9*. Like in the T-Score model, in the POMP score model, none of the teacher characteristics were predictors of the difference factor at Time 1 or Time 3.

   Final Model: Adding The Student Teacher Relationship. When the STRS LGM was added to the model, the larger model maintained good model fit, $\chi^2(37) = 52.32$, $p = .05$; RMSEA = .05, 90% CI = .00-.08; CFI = .98, TLI = .97. The relations of the latent STRS intercept and latent STRS slope with the latent difference factor intercept and slope are displayed in *Table 11*. Like in the T-score model, the latent STRS intercept had a positive relation with the latent difference factor intercept, .74, SE = .13, $z = 5.67$, $p < .001$, and a negative relation with the latent difference factor slope, -.35, SE = .09, $z = -3.74$, $p < .001$. This indicates that, at Time 1, the stronger the relationship between students and teachers, the more discrepant parents and teachers were in their ratings of externalizing problem behaviors. However, across time, stronger student-teacher relationships were related to decreased parent-teacher discrepancies in ratings of

externalizing behavior. Similar to the T-score model, the latent STRS slope had a negative relation with the latent different factor intercept in the POMP score model, -.36, SE = .11, z = -3.46, p = .001, and a positive relation with the latent difference factor slope, .41, SE = .08, z = 5.04, p < .001. Just as with the T-score model, this indicates that in the POMP score model, increases in the student-teacher relationship across time were associated with lower initial parent-teacher discrepancies in ratings of externalizing behavior at Time 1. Increases in the student-teacher relationship across time were also related to increases in parent-teacher discrepancies in ratings of externalizing behavior across time.

Results of the final model are displayed in *Figure 20.* Like in the T-score final model, child IQ ceased to be a predictor of higher parent-teacher discrepancies in internalizing behavior at Time 1 in the final model, .00, SE = .00, z = .85, p = .40. Higher parent income was still predictive of lower parent-teacher discrepancies in ratings of internalizing behavior at Time 1, -.007, SE = .002, z = -3.24, p = .001. Increasing child age was also predictive of a decrease in the difference factor intercept and an increase in the difference factor slope, Intercept on Age: -.002, SE = .001, z = -3.89, p < .001; Slope on Age: .001, SE = .00, z = 2.88, p = .004. In other words, increasing child age predicted a decrease in parent-teacher discrepancies in ratings of externalizing at Time 1 and an increase in the growth of parent-teacher discrepancies in externalizing behavior rating across time.

Discussion

The aims of this study were to analyze the change over time and across age in parent and teacher reports of internalizing and externalizing behavior problems, to examine parent-teacher discrepancies in ratings of child behavior longitudinally, and to determine child, parent, and teacher-characteristics which predict behavior rating discrepancies.

Examining Parent and Teacher Reports of Behavior Longitudinally

The first aim of the study was to examine the trajectories of child behavior problems, focusing on changes across child age. Only parent ratings showed a significant change across child age. Parent reports of internalizing and externalizing behavior increased with child age, but teacher reports of internalizing and externalizing problem behaviors demonstrated no change with increasing child age.

These results are consistent with other findings in the literature on typically developing (TD) children. Bongers, Koot, van der Ende, and Verhulst (2003) examined developmental trajectories of parent-reported child behavior problems across age in a representative sample of over two thousand typically developing youth ages 4 to 18. Consistent with the current study, they found that internalizing behavior problems increased with child age. It may be that TD children and children with ASD follow different trajectories for externalizing behavior. In TD youth, externalizing behavior problems decreased over time while the current study found an increase in child externalizing behavior for children with ASD.

The current study's finding that child behavior problems during the early transition to school tended to increase across age is also consistent with some of the ASD literature. With regard to internalizing problems, White, Oswald, Ollendick, and Scahill's (2009) review of the literature on comorbid anxiety and ASD found that, across studies, anxiety risk increased with age. When examining child behavior problems at ages 3, 5 and 7 years, Midouhas, Yogaratnam, Flouri, and Charman (2013) found that across problem types on the Strengths and Difficulties Questionnaire (i.e., emotional symptoms, conduct problems, hyperactivity/inattention, and peer relationship problems), child behavior problems increased across the age groups.

Parent-Teacher Discrepancies in Ratings of Child Behavior Across Time

The next aim of the study was to examine parent-teacher discrepancies in ratings of child behavior longitudinally and to determine the magnitude of the discrepancies. This study found that, on average, parents rated problem behaviors higher than teachers across problem types, a finding consistent with both the TD and ASD literatures (Berg-Nielsen et al., 2012; Kanne et al., 2009). Parent-teacher discrepancies were also greater for internalizing problems than for externalizing problems. This is consistent with many previous studies that have examined parent-teacher discrepancies in the TD (Achenbach et al., 1987; De los Reyes et al., 2015) and ASD literatures (Stratis & Lecavalier, 205; Kanne et al., 2009; Llanes, Blacher, Stavropoulos, & Eisenhower, 2018).

When examining informant discrepancies longitudinally, this study found no significant change over time in parent-teacher discrepancies in ratings of child internalizing or externalizing behavior. This finding is significant because few studies

have examined informant discrepancies longitudinally and no studies were found that examined discrepancies over time with a sample of children with ASD. Van Dulmen and Egeland (2011) examined trajectories of parent-teacher discrepancies in a sample of TD youth from age 7 to age 17.5. They found that an increase in parent-teacher discrepancies in ratings of internalizing and externalizing behavior over time. They found that teacher ratings of problem behaviors tended to decrease whereas parent ratings remained stable, leading to larger discrepancies across time. The findings of the current study were not consistent with Van Dulmen and Egeland's results, suggesting a difference between ASD and TD youth. Unlike discrepancies in ratings of TD youth, the current study indicated that, for children with ASD, parent-teacher discrepancies are stable across time.

Child, Parent, and Teacher Characteristics that Predict Discrepancies

The final aim of the study was to determine which child, parent, and teacher characteristics predict parent-teacher discrepancies in ratings of internalizing and externalizing behavior. Previous studies have identified child and parent characteristics that correlated with informant agreement, but no studies were found that examined the impact of teacher characteristics on parent-teacher rating discrepancies in children with ASD. This study aimed to explore this gap in the literature.

Child and Parent Characteristics. Child and parent characteristics that were predictive of parent-teacher discrepancies differed according to the type of analysis used. For internalizing behavior problems, parent income was predictive of parent-teacher discrepancies, with lower parent income predicting higher discrepancies in ratings of internalizing behavior at Time 1. Higher child IQ was also predictive of greater

discrepancies in parent-teacher ratings of internalizing behavior at Time 1, using T-scores as the unit of measurement. Finally, being female was predictive of an increase in the trajectory of parent-teacher discrepancies in ratings of internalizing behavior over time in the model using POMP scores. Child age was not a significant predictor of parent-teacher discrepancies in ratings of internalizing behavior. Thus, for internalizing behavior problems, lower parent income, higher child IQ, and being female predicted greater parent-teacher discrepancies.

For externalizing behavior, lower parent income also predicted higher parent-teacher discrepancies at Time 1. Child age predicted a decrease in parent-teacher discrepancies in ratings of externalizing behavior at Time 1 when using POMP scores. Child IQ and gender were not predictors of informant discrepancies in ratings of externalizing behavior. The finding regarding gender is inconsistent with findings in the TD literature. In their meta-analysis, De Los Reyes et al. (2015) found that informants had higher agreement when rating girls' behavior compared to boys; the opposite was found in this study. It is possible that for young children with ASD, girls may either be more difficult to rate than boys or they may demonstrate greater variability in their behavior across settings.

Previous studies in both TD and ASD literatures have also found that age and IQ are significant predictors of informant discrepancies. For TD youth and youth with ASD, informants tend to have higher agreement when rating younger children than adolescents (Achenbach et al., 1987; De Los Reyes & Kazdin, 2005; Schroeder et al., 2010; Stratis & Lecavalier, 2015). Van Dulmen and Egeland (2011) also found that parent-teacher

discrepancies in ratings of child externalizing behavior were smaller in younger children than adolescents. However, in the present study, the opposite was true; increasing age predicted a decrease in parent-teacher discrepancies in externalizing behavior. This may be because most of the literature examined differences in age between children and adolescents while the current study focused on early childhood, ages 4 to 7 years old.

In the ASD literature, IQ was found to be a moderator of informant discrepancies, such that with increasing IQ, agreement among raters on internalizing problem behaviors decreased (Stratis & Lecavalier, 2015). This is consistent with the current study's finding that higher IQ was predictive of greater parent-teacher discrepancies in internalizing behavior. Although the TD and ASD literature are mixed regarding the association between parent SES and informant discrepancies (Collishaw et al., 2009; De Los Reyes & Kazdin, 2005; Stone et al., 2013; Youngstrum et al., 2000; Zahner & Daskalakis, 1998), the current study found a consistent relation regarding parent income and parent-teacher discrepancies in both internalizing and externalizing behavior ratings. Lower parent income predicted larger parent-teacher discrepancies across behavior types and units of measurement.

Teacher Characteristics and the Student-Teacher Relationship. Whether teachers were in general or special education, autism training experience, class size and years of teaching experience were also examined in this study. Based on studies conducted with TD youth, it was hypothesized that these characteristics might impact a teacher's knowledge of the student with ASD and might thus lead them to rate child behavior more consistently with parents. In their study of parent-teacher agreement in

ratings of internalizing and externalizing behavior in TD youth, Zahner & Daskalakis (1998) found that the number of hours of weekly classroom teaching (a teacher characteristic that they believed represented teacher's familiarity with the student) significantly predicted agreement on both the internalizing and externalizing scales. However, in the current study, none of the teacher characteristics predicted discrepancies in parent-teacher ratings of either internalizing or externalizing problems.

The association between informant discrepancies and the student-teacher relationship was also examined. This study found that at Time 1, higher student-teacher relationship quality scores were related to higher parent-teacher discrepancies in ratings of both internalizing and externalizing behavior problems. In addition, for both problem types, increases in the trajectory of the student-teacher relationship across time was also associated with increases in the trajectory of parent-teacher discrepancies across time. These findings were surprising because it was hypothesized that higher student-teacher relationship quality would lead teachers to get to know their students better and therefore rate them more consistently with parents. Instead, a higher student-teacher relationship quality score was related to higher parent-teacher discrepancies. It may be that teachers who have good relationships with their students are able to manage problem behaviors in the classroom more effectively and, therefore, these teachers tend to rate fewer problem behaviors when compared to parents.

There is some support for this explanation from the TD literature. Although no studies have examined the impact of the student-teacher relationship on informant discrepancies in samples of youth with ASD, one study examined these relationships for

their impact on parent-teacher discrepancies in TD youth. Berg-Nielen and colleagues (2012) found that the student-teacher relationship was a significant predictor of informant disagreement between parents and teachers for ratings of internalizing problems. Specifically, when teachers perceived greater conflict in their relationship with the student, they tended to report more internalizing and externalizing problem behaviors.

T-Scores vs. POMP Scores as Unit of Measurement

Methodologically speaking, findings from this study support the use of either T-scores or POMP scores as units of measurement when analyzing informant discrepancies on the CBCL and TRF. Since the results of the models using these different scores were also very similar, either can be used, although T-scores are more easily interpretable for researchers and clinicians alike.

Limitations

This study used advanced statistical methods to examine informant discrepancies longitudinally, but it is not without its limitations. For example, this study was conducted using a sample of young children with ASD and the results of this study may not accurately reflect parameter estimates for informant discrepancies for a broader age range of school children. Another limitation is the fact that there were only three time points of measurement, spanning an 18ˉmonth period. The brief window of time in children's development may not be long enough to see changes in the trajectory of parent-teacher ratings across time. Although most children had a different teacher at Time 3, this design feature was built into the study and thus could be perceived as a strength.

Implications and Future Directions

The study of informant discrepancies has important implications for both research and practice. In clinical research, multi-informant reports of child outcomes are often used to support the efficacy of a treatment (Weisz, Jensen Doss, & Hawley, 2005). However, informants may report widely different information concerning child behavior. Exploring the reasons why informants may disagree (e.g., reporting setting-specific variation in behavior or rater-bias) becomes particularly important in considering the validity of treatment efficacy data.

In practice, multi-informant discrepancies can complicate treatment planning when children, parents, and teachers do not agree on a target problem. Parent-teacher rating discrepancies are particularly relevant in determining whether or not a child qualifies for school services. If parents and teachers do not agree, school professionals often default to teacher reports without considering factors that may influence rater discrepancies. Considerations of parent and teacher discrepancies in reports of behavioral symptoms are particularly important for children with ASD, who often have difficulty accurately reporting their own behavioral symptoms (Mazefsky, Kao, & Oswald, 2011). If parents and teachers are the only informants, it would be useful to weigh factors that may influence their agreement, such as child IQ, parent income, and teacher's relationship quality with the student, in order to ensure that the child's needs are met across the home and school environments. This study also determined that teachers tend to report lower levels of problem behavior than parents, particularly when rating internalizing behavior. Whether this discrepancy is due to rater bias or environmental

differences between home and school settings (or, likely, both), it is nonetheless important for school psychologists to ensure that parent concerns are heard and addressed to preserve an effective home-school partnership in treatment of the child.

An important finding from the current study is that the student-teacher relationship was significantly related to parent-teacher discrepancies in ratings of internalizing and externalizing behavior. Specifically, the more positive the relationship quality, the larger the parent-teacher discrepancies. For youth with ASD, this may be evidence that teachers who perceive themselves as having a stronger or more positive student-teacher relationship are reporting fewer problem behaviors in the classroom. This could be a manifestation of teacher perceptions of the child, or the fact that teachers might be implementing effective ways of dealing with child behavior in the classroom. Thus, the gap between parent and teacher ratings of behavior problems grows as parents tend to rate more behavior problems than teachers. This may underscore the need for continual parent training in behavior problem management.

Despite the above, the student-teacher relationship quality has been demonstrated to have a transactional relationship with externalizing behavior problems for TD youth. In other words, higher levels of externalizing problems predict poorer student-teacher relationship quality and poor relationship quality also predicts higher levels of externalizing problems in the future (Eisenhower, Baker, & Blacher, 2007; Hamre & Pianta, 2001). However, in youth with ASD, this appears to be a more child-driven pathway, where early teacher-reported externalizing behavior problems led to poorer student-teacher relationship quality over time (Eisenhower, Blacher, & Hurst Bush,

2015). This suggests that early teacher-reported child behavior problems should be addressed via behavioral intervention in the schools as quickly as possible in order to preserve good student-teacher relationships. Because student-teacher relationships are predictive of later student outcomes (Decker, Dona, & Christenson, 2007; Hamre & Pianta, 2001; Pianta & Stuhlman, 2004), it is important that early teacher concerns about behavior problems in students with ASD be addressed quickly.

This study examined multi-informant behavior problems in preschool and early school-aged children with ASD. Studying this population is critical to understanding the challenges faced by youth and their families as they transition to school. Even typically developing children who begin school in kindergarten and the early grades experience new academic, social, emotional, and behavioral demands (Rimm-Kaufman & Pianta, 2000; Pianta, Cox, Taylor, & Early, 1999). The transition to school can be even more challenging for children with ASD, who are at risk for school problems, disruptive behavior problems, and social rejection (Ashburner, Ziviani, & Rodger, 2010). This study revealed that internalizing and externalizing behavior problems did increase with increasing child age. Knowledge of these trajectories may be useful for examining critical junctures for intervention, in order to improve long-term positive school outcomes, particularly when it comes to addressing child internalizing problems.

Finally, researchers and clinicians alike need to understand the ways in which rater characteristics and environmental differences impact the reports of parent and teachers in order to accurately assess and effectively treat youth with ASD. Future work should further examine teacher characteristics, such as how long the teacher has known

72

the student and classroom practices, which may impact informant discrepancies. Researchers should also further examine informant discrepancies over greater periods of time than this study allowed. Understanding how and why parents and teachers disagree is an important step to coordinating treatment and creating strong home-school partnerships.

References

Achenbach, T.M. (1995). Empirically based assessment and taxonomy: Applications to clinical research. *Psychological Assessment, 7*, 261-274.

Achenbach, T. M. (2011). Commentary: Definitely more than measurement error: But how should we understand and deal with informant discrepancies? *Journal of Clinical Child & Adolescent Psychology, 40*, 80-86.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213.

Achenbach, T.M., & Rescorla, L.A. (2001). *Manual for the ASEBA school-age forms and profiles: An integrated system of multi-informant assessment*. Burlington, VT: Research Center for Children, Youth, and Families.

Ashburner, J., Ziviani, J., & Rodger, S. (2010). Surviving in the mainstream: Capacity of children with autism spectrum disorders to perform academically and regulate their emotions and behavior at school. *Research in Autism Spectrum Disorders, 4*, 18-27.

Baker, B. L., & Blacher, J. (2015). Disruptive behavior disorders in adolescents with ASD: Comparison to youth with intellectual disability or typical cognitive development. *Journal of Mental Health Research in Intellectual Disabilities, 8*, 98-116.

Barnett, R. C, Marshall, N. L., Raudenbush, S. W., & Brennan, R. T. (1993). Gender and The relationship between job experiences and psychological distress: A study of dual-earner couples. *Journal of Personality and Social Psychology, 64*, 794-806.

Bauminger, N., Solomon, M., & Rogers, S.J. (2010). Externalizing and internalizing behaviors in ASD. *Autism Research, 3*, 101-112.

Berg-Nielsen, T., Solheim, E., Belsky, J., & Wichstrom, L. (2012). Preschoolers' psychosocial problems: In the eye of the beholder? Adding teacher characteristics as determinants of discrepant parent-teacher reports. *Child Psychiatry and Human Development, 43*, 393-413.

Biglan, A. *Changing culture practices: A contextualistic framework for intervention research*. Reno: Context Press, 1995.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*, 371-399.

Brereton, A.V., Tonge, B.J., & Einfeld, S.L. (2006). Psychopathology in children and adolescents with autism compared to young people with intellectual disability. *Journal of Autism and Developmental Disorders, 36*, 863-870.

Bongers, I.L., Koot, H.M., van der End, J., & Verhulst, F.C. (2003). The normative development of child and adolescent problem behaviors. *Journal of Abnormal Psychology, 112*(2), 179-192.

Byrne, B.M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York, NY: Routledge.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Cichetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284-290.

Cohen, P., Cohen J., Aiken, L.S., & West, S.G. (1999). The problem of units and the circumstance for POMP. *Multivariate Behavioral Research, 34*(3), 315-346.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation for the behavioral sciences* (3rd ed.) Mahwah: Erlbaum.

Collishaw, S., Goodman, R., Ford, T., Rabe-Hesketh, S., & Pickles, A. (2009). How far are associations between child, family, and community factors and child psychopathology informant-specific and informant-general? *Journal of Child Psychology and Psychiatry, 50*, 571-580.

Decker, D. M., Dona, D. P., & Christenson, S. L. (2007). Behaviorally at-risk African American students: The importance of student–teacher relationships for student outcomes. *Journal of School Psychology, 45*(1), 83-109.

De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology, 40*, 1-9.

De Los Reyes, A., Alfano, C. A., & Beidel, D. C. (2010). The relations among measurements of informant discrepancies within a multisite trial of treatments for childhood social phobia. *Journal of Abnormal Child Psychology, 38*, 395-404.

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. Psychological Bulletin, 141, 858-900.

De Los Reyes, A., Ehrlich, K.B., Swan, A.J., Luo, T.J., Van Wie, M., & Pabon, S.C. (2013). An experimental test of whether informants can report about child and family behavior based on settings of behavioral expression. *Journal of Child and Family Studies, 22*, 177-191.

De Los Reyes, A., Goodman, K. L., Kliewer, W., & Reid-Quiñones, K. (2008). Whose depression relates to discrepancies? Testing relations between informant characteristics and informant discrepancies from both informants' perspectives. *Psychological Assessment, 20*, 139-149.

De Los Reyes, A., Henry, D.B., Tolan, P.H., & Wakschlag, L.S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology, 37*, 637-652.

De Los Reyes, A., & Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment, 16*, 330-334.

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*, 483.

De Los Reyes, A., Thomas, S.A., Goodman, K.L., & Kundey, S.M.A. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology, 9*, 123-149.

Dodge, K. A., Pettit, G. S., & Bates, J. E. (1994). Socialization mediators of the relation between socioeconomic status and child conduct problems. *Child Development, 65*, 649-665.

Dumenci, L., Achenbach, T.M., & Windle, M. (2011). Measuring context-specific and cross contextual components of hierarchical constructs. *Journal of Psychopathology and Behavioral Assessment, 33*, 3-10.

Eisenhower, A.S., Baker, B.L., & Blacher, J. (2007). Early student-teacher relationships of children with and without intellectual disability: Contributions of behavioral, social, and self-regulatory competence. *Journal of School Psychology, 45*, 363-383.

Eisenhower, A.S., Blacher, J., & Hurst Bush, H. (2015). Longitudinal associations between externalizing problems and student-teacher relationship quality for young children with ASD. *Research in Autism Spectrum Disorders, 9*, 163-173.

Edwards, J.R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes, 58*, 51-100.

Ferdinand, R. F., Van Der Ende, J., & Verhulst, F. C. (2007). Parent–teacher disagreement regarding psychopathology in children: A risk factor for adverse outcome? *Acta Psychiatrica Scandinavica, 115*, 48-55.

Gillberg, C., & Fernell, E. (2014). Autism plus versus autism pure. *Journal of Autism and Developmental Disorders, 44*, 3274-3276.

Green, S. A., Berkovits, L. D., & Baker, B. L. (2015). Symptoms and development of anxiety in children with or without intellectual disability. *Journal of Clinical Child & Adolescent Psychology, 44*(1), 137-144.

Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010). Cross informant agreement for ratings for social skill and problem behavior ratings: An investigation of the Social Skills Improvement System—Rating Scales. *Psychological Assessment, 22*, 157-166.

Grietens, H., Onghena, P., Prinzie, P., Gadeyne, E., Van Assche, V., Ghesquière, P., & Hellinckx, W. (2004). Comparison of mothers', fathers', and teachers' reports on problem behavior in 5-to 6-year-old children. *Journal of Psychopathology and Behavioral Assessment, 26*, 137-146.

Hamre, B.K., & Pianta, R.C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development, 72*, 625-638.

Hattier, M. A., Matson, J. L., Belva, B. C., & Adams, H. (2013). The effects of informant age and education level on childhood ASD symptom endorsement. *Research in Autism Spectrum Disorders, 7*, 398-404.

Hartley, A.G., Zakriski, A.L., & Wright, J.C. (2011). Probing the depths of informant discrepancies: Contextual influences on divergence and convergence. *Journal of Clinical Child and Adolescent Psychology, 40*, 54-66.

IBM Corp. Released 2016. *IBM SPSS Statistics for Windows*, Version 24.0. Armonk, NY: IBM Corp.

Kanne, S.M., Abbacchi, A.M., & Constantino, J.N. (2009). Multi-informant ratings of psychiatric symptom severity in children with autism spectrum disorders: The importance of environmental context. *Journal of Autism and Developmental Disorders, 39*, 856-864.

Kim, J.A., Szatmari, P., Bryson, S.E., Streiner, D.L., & Wilson, F.J. (2000). The prevalence of anxiety and mood problems among children with autism and Asperger syndrome. *Autism, 4*, 117-132.

Kraemer, H. C., Measell, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry, 160*, 1566-1577.

Jones, E.E., & Nesbett, R.E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. E.E. Jones, D.E., Kanouse, H.H. Kelly, R.E. Nispett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (79-94). Morristown: General Learning Press.

Laird, R.D., & De Los Reyes, A. (2013). Testing informant discrepancies as predictors of early adolescent psychopathology: Why difference scores cannot tell you what you want to know and how polynomial regression may. *Journal of Abnormal Child Psychology, 41*, 1-14.

Laird, R.D., & Weems, C.F. (2011). The equivalence of regression models using difference scores and models using separate scores from each informant: Implications for the study of informant discrepancies. *Psychological Assessment, 23*, 388-397.

Leyfer, O.T., Folstein, S.E., Bacalman, S., Davis, N.O., Dinh, E., Morgan, J., Tager Flusberg, H., & Lainhart, J.E. (2006). Comorbid psychiatric disorders in children with autism: Interview development and rates of disorders. *Journal of Autism and Developmental Disorders, 36*, 849-861.

Little, R., & Rubin, D. (2014). *Statistical analysis with missing data*. New York, NY: John Wiley & Songs, Inc.

Llanes, E., Blacher, J., Stavropoulos, K., & Eisenhower, A. (2018). Parent and teacher reports of comorbid anxiety and ADHD symptoms in children with ASD. *Journal of autism and developmental disorders*, 1-12.

Lord, C., Rutter, M., Dilavore, P. and Risi, S. (2008) *Autism Diagnostic Observation Schedule Manual*. Los Angeles, CA: Western Psychological Services.

Lord, F.M. (1963) Elementary models for measuring change. In. C.W. Harris (Eds.), *Problems measuring change*. Madison, WI: University of Wisconsin Press.

Magnuson, K.M., & Constantino, J.N. (2011). Characterization of depression in children with autism spectrum disorders. *Journal of Developmental and Behavioral Pediatrics, 32*, 332-340.

Mattila, M.L., Hurtig, T., Haapsamo, H., Jussila, K., Kuusikko-Gauffin, S., Kielinen, M., Linna, S.L., Ebeling, H., Bloigu, R., Joskitt, L., Pauls, D.L., & Moilanen, I. (2010). Comorbid psychiatric disorders associated with Asperger syndrome/high functioning autism: A community and clinic-based study. *Journal of Autism and Developmental Disorders, 40*, 1080-1093.

Mazefsky, C. A., Kao, J., & Oswald, D. P. (2011). Preliminary evidence suggesting caution in the use of psychiatric self-report measures with adolescents with high functioning autism spectrum disorders. *Research in Autism Spectrum Disorders, 5*, 164-174.

McArdle, J.J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60*, 577-605.

Mehta, P.D., & West, S.G. (2000). Putting the individual back into individual growth curves. *Psychological Methods, 5*, 23-43.

Merrell, K. W. (2007). *Behavioral, social and emotional assessment of children and adolescents* (3rd ed.). New York, New York: Lawrence Erlbaum Associates.

Midouhas, E., Yogaratnam, A., Flouri, E., & Charman, T. (2013). Psychopathology trajectories of children with autism spectrum disorder: The role of family poverty and parenting. *Journal of the American Academy of Child and Adolescent Psychiatry, 52*(10), 1057-1065.

Morris, E.K., (1988). Contextualism: The world view of behavior analysis. *Journal of Experimental Child Psychology, 46*, 289-323.

Neece, C.L., Green, S.A., & Baker, B.L. (2012). Parenting stress and child behavior problems: A transactional relationship across time. *American Journal for Intellectual and Developmental Disabilities, 117*(1), 48-66.

Pianta, R.C. (2001). *Student–teacher relationship scale–short form*. Lutz, FL: Psychological Assessment Resources.

Pianta, R. C., Cox, M. J., Taylor, L., & Early, D. (1999). Kindergarten teachers' practices related to the transition to school: Results of a national survey. *The Elementary School Journal, 100*(1), 71-86.

Pianta, R. C., & Stuhlman, M. W. (2004). Teacher-child relationships and children's success in the first years of school. *School Psychology Review, 33*(3), 444-458.

Raudenbush, S.W., Brennan, R.T., & Barnett, R.C. (1995). A multivariate hierarchical model for studying psychological change within married couples. *Journal of Family Psychology, 9*, 161-174.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage Publications.

Reed, P., & Osborne, L. A. (2013). The role of parenting stress in discrepancies between parent and teacher ratings of behavior problems in young children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*, 471-477.

Richters, J. E. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. *Psychological Bulletin, 112*, 485–499.

Richters, J., & Pellegrini, D. (1989). Depressed mothers' judgments about their children: An examination of the depression-distortion hypothesis. *Child Development, 60*, 1068-1075.

Rimm-Kaufman, S. E., Pianta, R. C., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly, 15*(2), 147-166.

Rogosa, D.R., & Willett, J.B. (1983). Demonstrating the reliability of the difference score in the measurement of change, Journal of Educational Measurement, 20, 335-343.

Salbach-Andrae, H., Lenz, K., & Lehmkuhl, U. (2009). Patterns of agreement among parent, teacher, and young ratings in a referred sample. European Psychiatry, 24, 345-351.

Sattler, J., & Dumont, R. (2004). *Assessment of Children: WISC-IV and WPPSI Supplement*. San Diego, CA: Jerome M. Sattler Publisher, Inc.

Schroeder, J. F., Hood, M. M., & Hughes, H. M. (2010). Inter-parent agreement on the syndrome scales of the Child Behavior Checklist (CBCL): Correspondence and discrepancies. *Journal of Child & Family Studies, 19*, 646-654.

Simonoff, E., Pickles, A., Chaman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*, 921-929.

Stanger, C., & Lewis, M. (1993). Agreement among parents, teachers, and children on internalizing and externalizing behavior problems. *Journal of Clinical Child Psychology, 22*, 107-116.

Stratis, E.A., Lecavalier, L. (2015). Informant agreement for youth with autism spectrum disorder or intellectual disability: A meta-analysis. *Journal of Autism and Developmental Disorders, 4*, 1026-1041

Stone, S. L., Speltz, M. L., Collett, B., & Werler, M. M. (2013). Socioeconomic factors in relation to discrepancy in parent versus teacher ratings of child behavior. *Journal of Psychopathology and Behavioral Assessment, 35*, 314-320.

Tversky, B., & Marsh, E.J. (2000). Biased retellings of events yield biased memories. *Cognitive Psychology, 40*, 1-38.

Van der Ende, J., Verhulst, F. C., & Tiemeier, H. (2012). Agreement of informants on emotional and behavioral problems from childhood to adulthood. *Psychological Assessment, 24*, 293-300.

Van der Meer, J.M.J., Oerlemans, A.M., van Steijn, D.J., Lappenschaar, M.G.A., de Sonneville, L.M.J., Buitelaar, J.K., & Rommelse, N.N.J. (2012). Are autism spectrum disorder and attention-deficit/hyperactivity disorder different manifestations of one overarching disorder? Cognitive and symptom evidence from a clinical and population-based sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*, 1160-1172.

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*, 84-92.

Van Steensel, F.J.A., Bogels, S.M., & Perrin, S. (2011). Anxiety disorders in children and adolescents with autistic spectrum disorders: A meta-analysis. *Clinical Child and Family Psychology Review, 14*, 302-317.

Weisz, J. R., Jensen Doss, A., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology, 56*, 337–363.

Wechsler, D. (2002). *Manual for the Wechsler Intelligence Scale for Children—Third Edition*. San Antonio, TX: Psychological Corporation.

White, S.W., Oswald, D., Ollendick, T., & Scahill, L. (2009). Anxiety in children and adolescents with autism spectrum disorders. *Clinical Psychology Review, 29*(3), 216-229.

Youngstrum, E.A., Findling, R.L., & Calabrese, J.R. (2003). Who are the comorbid adolescents? Agreement between psychiatric diagnosis, youth, parent, and teacher report. *Journal of Abnormal Child Psychology, 31*, 231-245.

Zahner, G., & Daskalakis, C. (1998). Modelling sources of informant variance in parent and teacher ratings of child psychopathology. *International Journal of Methods in Psychiatric Research, 7*, 3-16.

Table 1.
Participant Demographics (*N* = 184)

Variable	Total sample *M (SD)*
Child Characteristics	
Child Age in Months at Time 1	67.6 (12.1)
Child IQ, *M* (*SD*)	87.8 (17.7)
Child Gender, % male (*n*)	81.5 (150)
Child ethnicity, % (*n*)	
African American	3.1 (5)
Asian	6.2 (10)
Caucasian	56.5 (91)
Latino	9.3 (15)
Other	24.8 (40)
Parent Characteristics	
Mothers, % (n)	87.0 (160)
Family income ≥ $65,000, % (*n*)	56.4 (101)
Mother's Education, % college degree or higher (*n*)	63.0 (116)
Teacher Characteristics	
Gender, % female (*n*)	
Teacher at Time 1 & 2	88.7 (133)
Teacher at Time 3	90.3 (112)
General Education teacher, % (*n*)	
Teacher at Time 1 & 2	55.4 (82)
Teacher at Time 3	62.9 (78)
Number of Years Teaching	
Teacher at Time 1 & 2	14.1 (9.2)
Teacher at Time 3	13.4 (9.0)

Table 2.
Number of Participants Across Time

Participant	Time Point	N
Parent	Time 1	184
	Time 2	180
	Time 3	163
Teacher	Time 1	145
	Time 2	142
	Time 3	163

Table 3.
Results of LGMs for Parent-reported Internalizing and Externalizing Behavior

Model	Intercept	Intercept Variance	Slope	Slope Variance	Slope with Intercept
Internalizing (T-score)	63.32*** (.72)	73.83*** (10.37)	-.94*** (.28)	.51 (1.07)	.15 (2.39)
Internalizing (POMP scores)	.25*** (.01)	.02*** (.02)	-.02*** (.003)	.00 (.00)	.001* (.001)
Externalizing (T-score)	60.39*** (.81)	103.58*** (12.60)	-.44* (.21)	2.76*** (.96)	-6.03* (2.62)
Externalizing (POMP scores)	.35*** (.02)	.04*** (.01)	-.03*** (.004)	.001* (.00)	-.004** (.001)

Note. Tabled values are parameter estimates, with standard errors in parentheses
*p<.05, **p<.01, ***p<.001

Table 4.
Results of LGMs for Teacher-reported Internalizing and Externalizing Behavior

Model	Intercept	Intercept Variance	Slope	Slope Variance	Slope with Intercept
Internalizing (T-score)	57.80*** (.90)	94.54*** (15.00)	-.71* (.38)	11.39*** (2.87)	-26.58*** (5.71)
Internalizing (POMP scores)	.17*** (.01)	.01*** (.002)	-.02*** (.01)	.002*** (.00)	-.004*** (.001)
Externalizing (T-score)	58.13*** (.76)	80.72*** (11.54)	-.78** (.31)	7.14*** (1.85)	-17.24*** (4.20)
Externalizing (POMP scores)	.23*** (.02)	.03*** (.004)	-.03*** (.01)	.002*** (.001)	-.006** (.001)

Note. Tabled values are parameter estimates, with standard errors in parentheses
*p<.05, **p<.01, ***p<.001

Table 5.
Results of LGMs for Parent- and Teacher-Reported Internalizing and Externalizing
Behavior across Child Age in Months

Model	Parent-Report		Teacher-Report	
	Intercept on Age	Slope on Age	Intercept on Age	Slope on Age
Internalizing (T-score)	-.20*** (.06)	.05** (.02)	-.09 (.08)	.03 (.03)
Internalizing (POMP scores)	-.004** (.001)	.001*** (.00)	-.003** (.001)	.00 (.00)
Externalizing (T-score)	-.01 (.07)	.01 (.02)	.03 (.06)	-.01 (.03)
Externalizing (POMP scores)	-.01*** (.001)	.001** (.00)	-.003* (.001)	.00 (.00)

Note. Tabled values are parameter estimates, with standard errors in parentheses
*p<.05, **p<.01, ***p<.001

Table 6.
Latent Growth Modeling Results for Similarity Factor

Model	Intercept	Intercept Variance	Slope	Slope Variance
Internalizing (T-score)	60.51*** (.58)	39.84*** (6.59)	-1.58*** (.45)	10.94** (3.61)
Internalizing (POMP scores)	.21*** (.01)	.009*** (.001)	-.035*** (.01)	.002* (.00)
Externalizing (T-score)	59.08*** (.62)	56.87*** (7.50)	-1.01** (.37)	7.75** (2.70)
Externalizing (POMP scores)	.29*** (.01)	.02*** (.003)	-.06*** (.01)	.003* (.001)

Note. Tabled values are parameter estimates, with standard errors in parentheses
*p<.05, **p<.01, ***p<.001

Table 7.
Latent Growth Modeling Results for Difference Factor

Model	Intercept	Intercept Variance	Slope	Slope Variance
Internalizing (T-score)	2.80*** (.57)	44.32*** (6.07)	-.30 (.42)	14.59*** (3.22)
Internalizing (POMP scores)	.04*** (.01)	.008*** (.001)	-.001 (.01)	.002*** (.00)
Externalizing (T-score)	1.31** (.49)	31.42*** (4.48)	.14 (.37)	10.54*** (2.66)
Externalizing (POMP scores)	.06*** (.01)	.01*** (.002)	-.01 (.01)	.003*** (.001)

Note. Tabled values are parameter estimates, with standard errors in parentheses
*p<.05, **p<.01, ***p<.001

Table 8.
Child and Parent Predictors of Latent Difference Factor Intercept and Slope

Model	Step	Predictor	Effect on Diff Factor Intercept	SE	Effect on Diff Factor Slope	SE
Int. T-score Model	Step 1: Adding Child Predictors	Child IQ	.24***	.06	-.10*	.05
		Child Age	-.10	.09	.03	.07
		Child Gender	.46	2.85	2.92	2.20
	Step 2: Adding Parent Predictors	Child IQ	.16***	.05	--	--
		Par Income	-.79*	.37	.20	.29
Int. POMP Score Model	Step 1: Adding Child Predictors	Child IQ	.003*	.001	-.001*	.001
		Child Age	-.002	.001	.00	.001
		Child Gender	.01	.04	.05*	.03
	Step 2: Adding Parent Predictors	Child IQ	.003***	.001	-.001*	.001
		Child Gender	--	--	.05**	.02
		Par Income	-.01*	.01	.001	.003
Ext. T-score Model	Step 1: Adding Child Predictors	Child IQ	.18**	.05	-.06	.04
		Child Age	-.05	.08	.05	.06
		Child Gender	-.48	2.48	.34	1.98
	Step 2: Adding Parent Predictors	Child IQ	.15**	.04	--	--
		Par Income	-.62*	.32	.04	.25
Ext. POMP Score Model	Step 1: Adding Child Predictors	Child IQ	.003**	.001	-.001	.001
		Child Age	-	.001	.002*	.001
		Child Gender	.006*** .03	.04	.04	.04

90

Step 2:	Child IQ	.002**	.001	--	--
Adding	Child Age	-	.001	.003**	.001
Parent	Par Income	.006***	.01	-.001	.004
Predictors		-.01*			

*p<.05, **p<.01, ***p<.001

Table 9.
Teacher Predictors of Difference Factor at Time 1 and 3

Model	Step	Predictor	Effect on Diff Factor Time 1	SE	Effect on Diff Factor Time 3	SE
Int. T-score Model	Step 3: Adding Teacher Predictors	Autism Exp Years Teaching Gen Ed/Sped Class size	-6.40 0.07 1.11 -.27	3.61 .14 3.55 .18	1.91 .09 1.85 .09	3.75 .15 3.66 .19
Int. POMP Score Model	Step 3: Adding Teacher Predictors	Autism Exp Years Teaching Gen Ed/Sped Class size	-.05 .002 -.04 -.002	.04 .002 .04 .002	.06 .001 .06 .000	.04 .002 .04 .002
Ext. T-score Model	Step 3: Adding Teacher Predictors	Autism Exp Years Teaching Gen Ed/Sped Class size	.70 .09 -3.60 -.06	3.28 .13 3.09 .16	-1.29 .04 1.75 .10	3.43 .15 3.35 .18
Ext. POMP Score Model	Step 3: Adding Teacher Predictors	Autism Exp Years Teaching Gen Ed/Sped Class size	-.04 .001 -.10 -.001	.06 .002 .05 .003	.03 .002 .02 .000	.05 .002 .05 .003

*p<.05, **p<.01, ***p<.001

Table 10.
Latent Growth Modeling Results of Student-Teacher Relationship Across Time

Model	Intercept	Intercept Variance	Slope	Slope Variance
Student-Teacher Relationship	109.14*** (1.05)	99.76*** (21.46)	1.27 (.90)	24.22*** (16.93)

Note. Tabled values are parameter estimates, with standard errors in parentheses
*p<.05, **p<.01, ***p<.001

Table 11.
Student Teacher Relationship Intercept and Slope Correlations with Difference Factor
Intercept and Slope

Model	Step		Difference Factor Int	Difference Factor Slope
Int. T-score Model	<u>Step 4</u>: Adding STRS LGM	STRS Int	45.78*** (8.40)	-22.74*** (6.00)
		STRS Slope	-17.19* (6.71)	17.20** (5.00)
Int. POMP Score Model	<u>Step 4</u>: Adding STRS LGM	STRS Int	0.51*** (.11)	-.25*** (.07)
		STRS Slope	-.18* (.09)	.18** (.06)
Ext. T-score Model	<u>Step 4</u>: Adding STRS LGM	STRS Int	38.42*** (7.25)	-17.07** (5.23)
		STRS Slope	-20.31** (5.98)	21.15*** (4.60)
Ext. POMP Score Model	<u>Step 4</u>: Adding STRS LGM	STRS Int	.74*** (.13)	-.35*** (.09)
		STRS Slope	-.36** (.11)	.41*** (.08)

Note. Tabled values are parameter estimates, with standard errors in parentheses
*$p<.05$, **$p<.01$, ***$p<.001$

Figure 1.
Proposed latent growth model examining parent-reported internalizing behavior across time
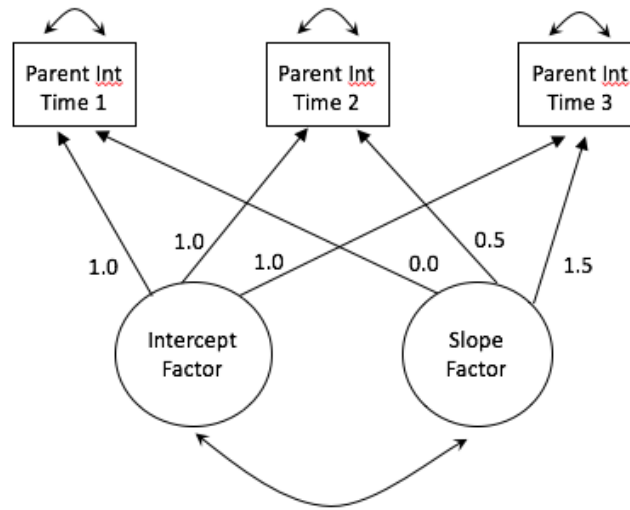
Figure 2.
Proposed latent growth model examining parent-reported externalizing behavior across time
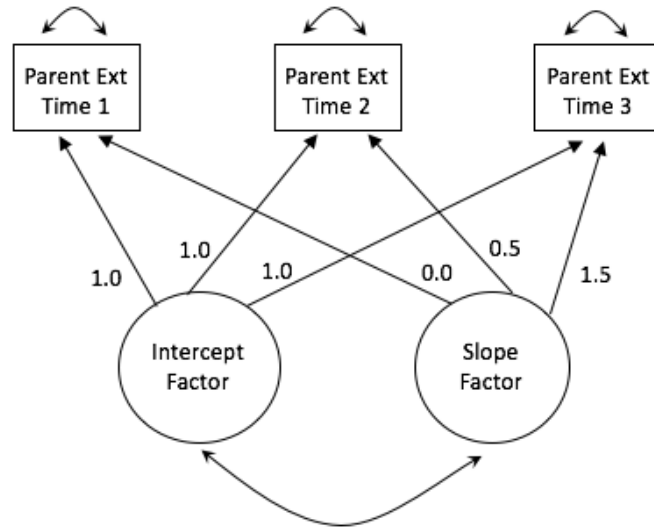
Figure 3.
Proposed latent growth model examining teacher-reported internalizing behavior across time
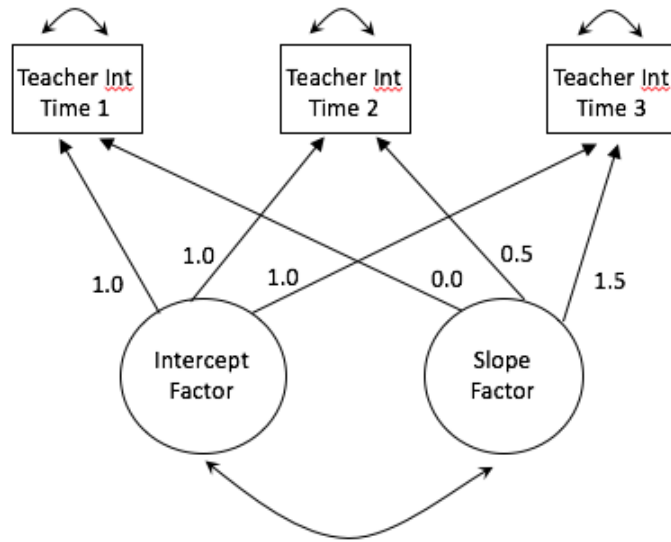
Figure 4.
Proposed latent growth model examining teacher-reported externalizing behavior across
time
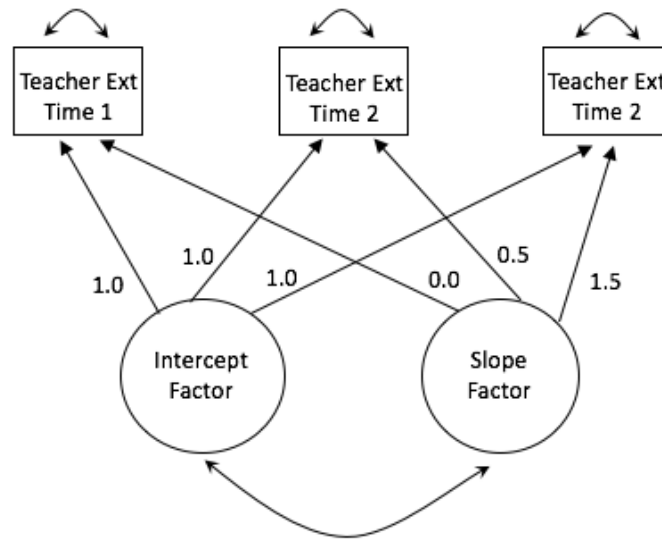
Figure 5.
Proposed latent growth model examining parent-reported internalizing behavior across age
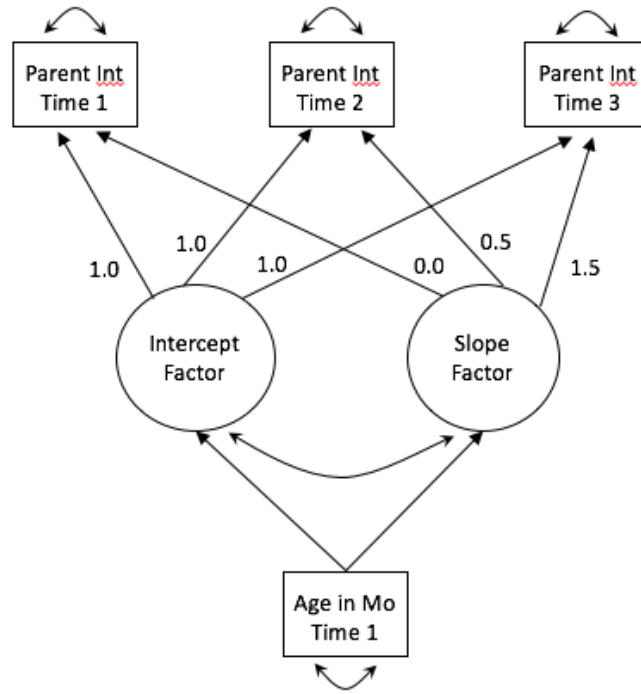
Figure 6.
Proposed latent growth model examining parent-reported externalizing behavior across age
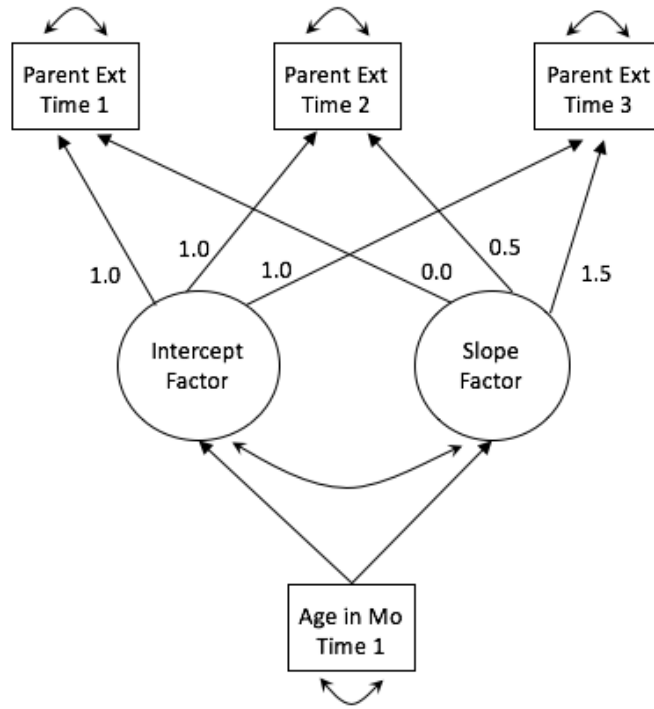
Figure 7.
Proposed latent growth model examining teacher-reported internalizing behavior across age
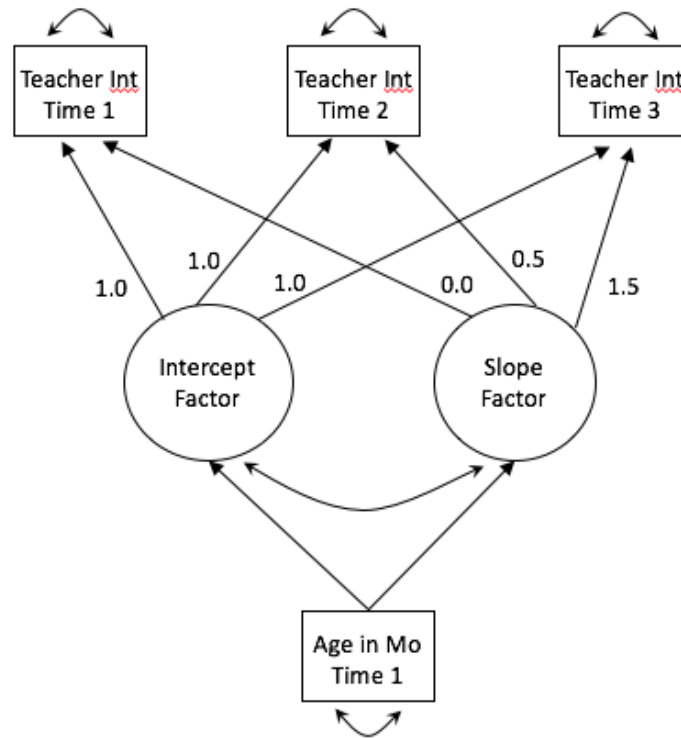
Figure 8.
Proposed latent growth model examining teacher-reported externalizing behavior across age
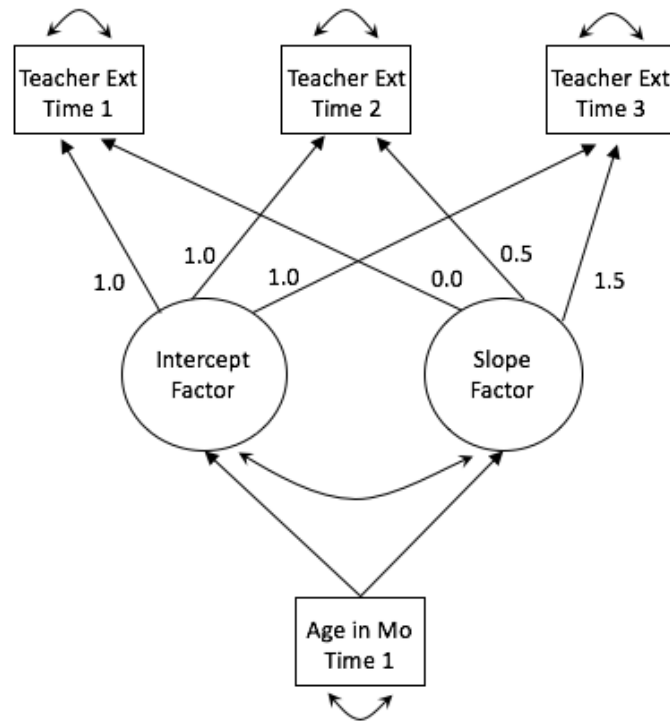
Figure 9.
Proposed model examining latent similarities and differences between parent and teacher-reported internalizing behavior at Time 1
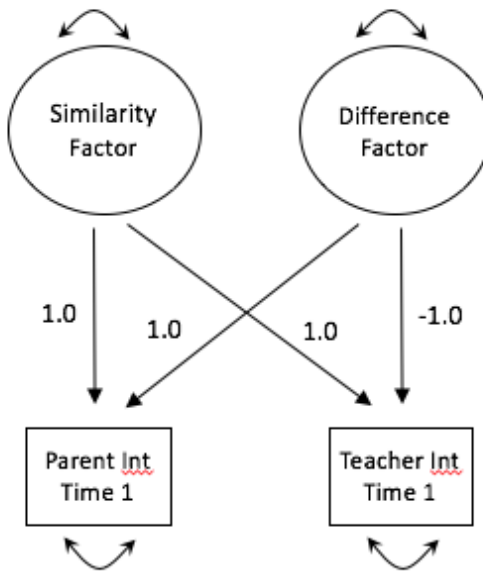
Figure 10.
Proposed model examining latent similarities and differences between parent and teacher-reported externalizing behavior at Time 1
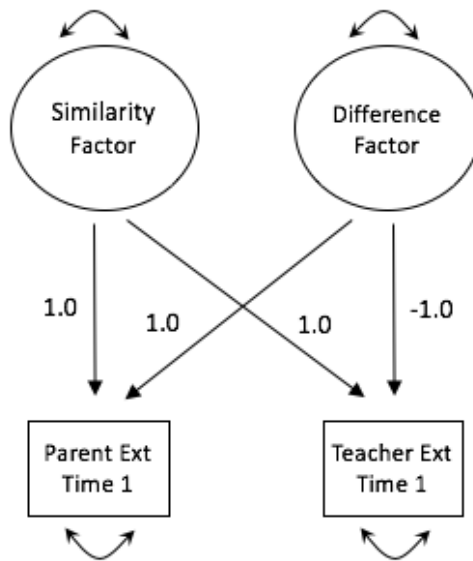
Figure 11.
Proposed model examining similarities between parent and teacher-reported internalizing behavior across time points
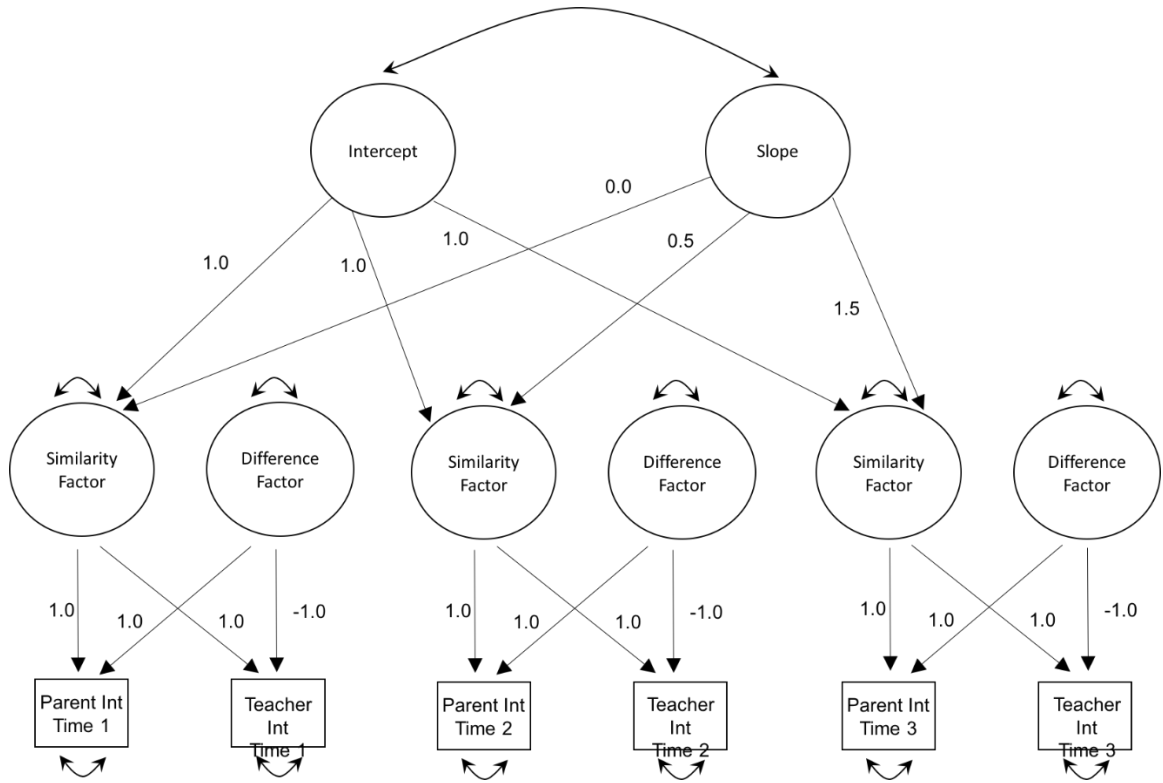
Figure 12.
Proposed model examining discrepancies in parent and teacher-reported internalizing
behavior across time points
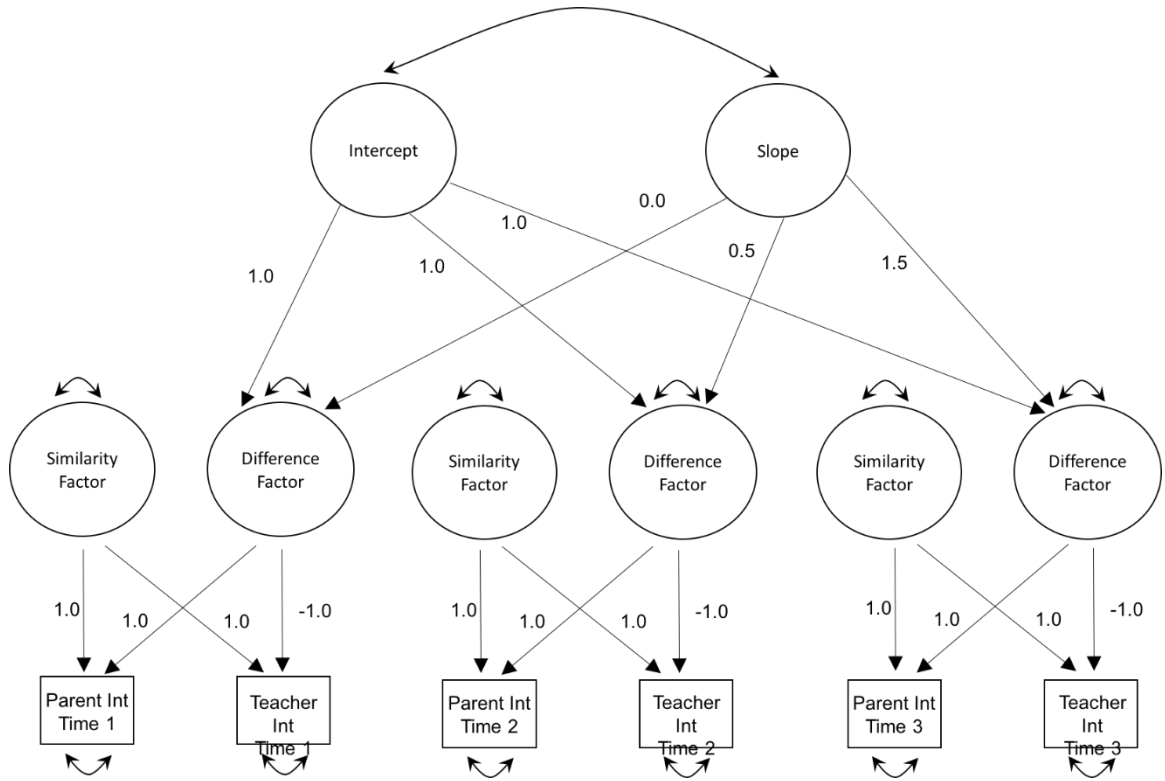
Figure 13.
Proposed model examining similarities between parent and teacher-reported externalizing behavior across time points
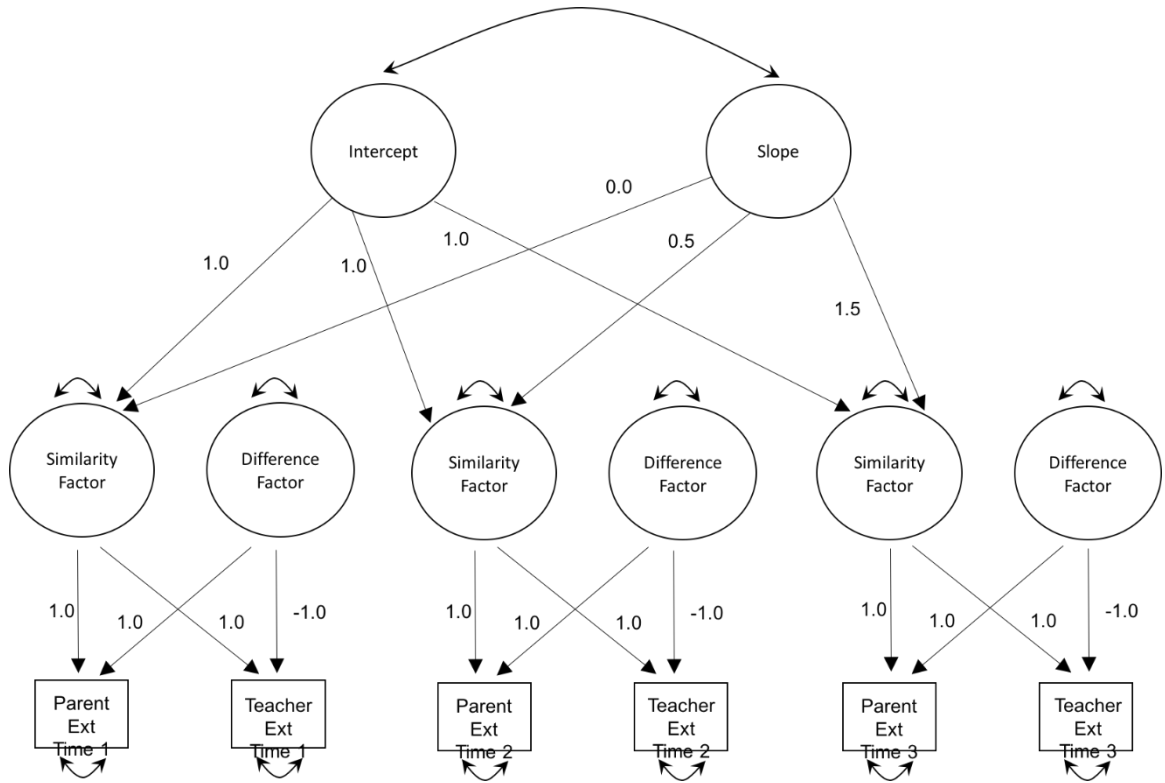
Figure 14.
Proposed model examining discrepancies in parent and teacher-reported externalizing
behavior across time points
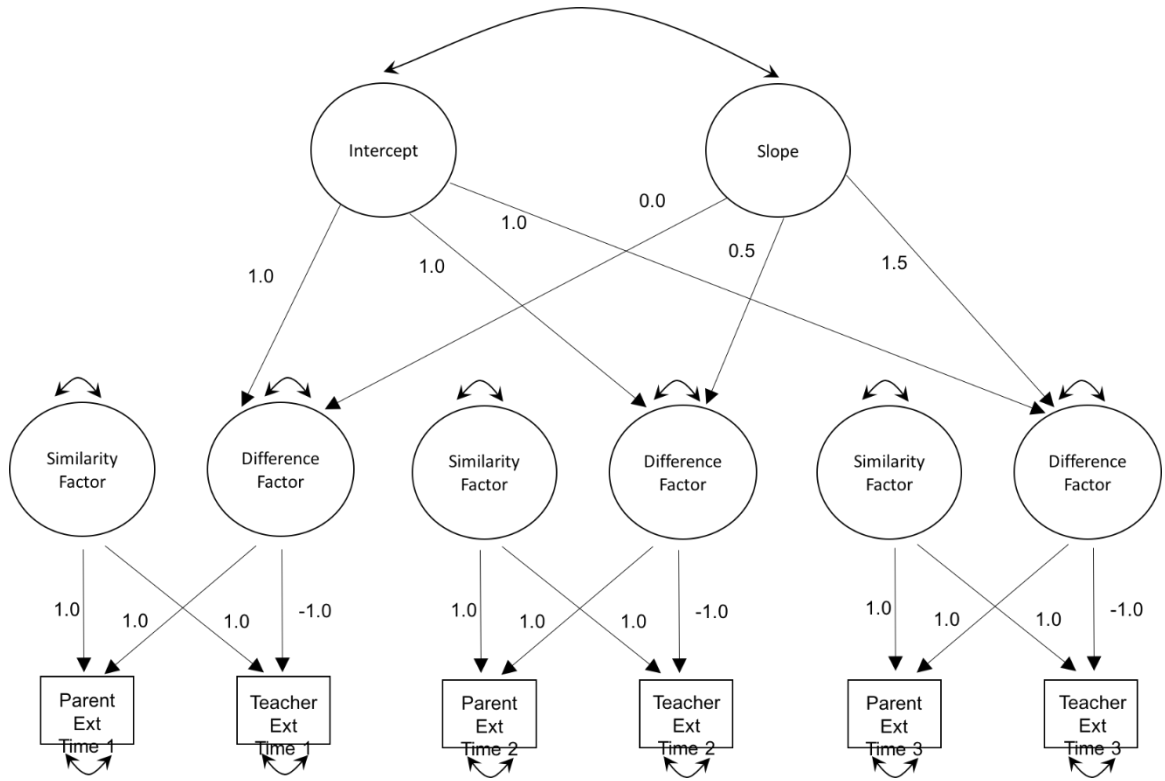
Figure 15.
Proposed model examining parent, child, and teacher characteristics that predict
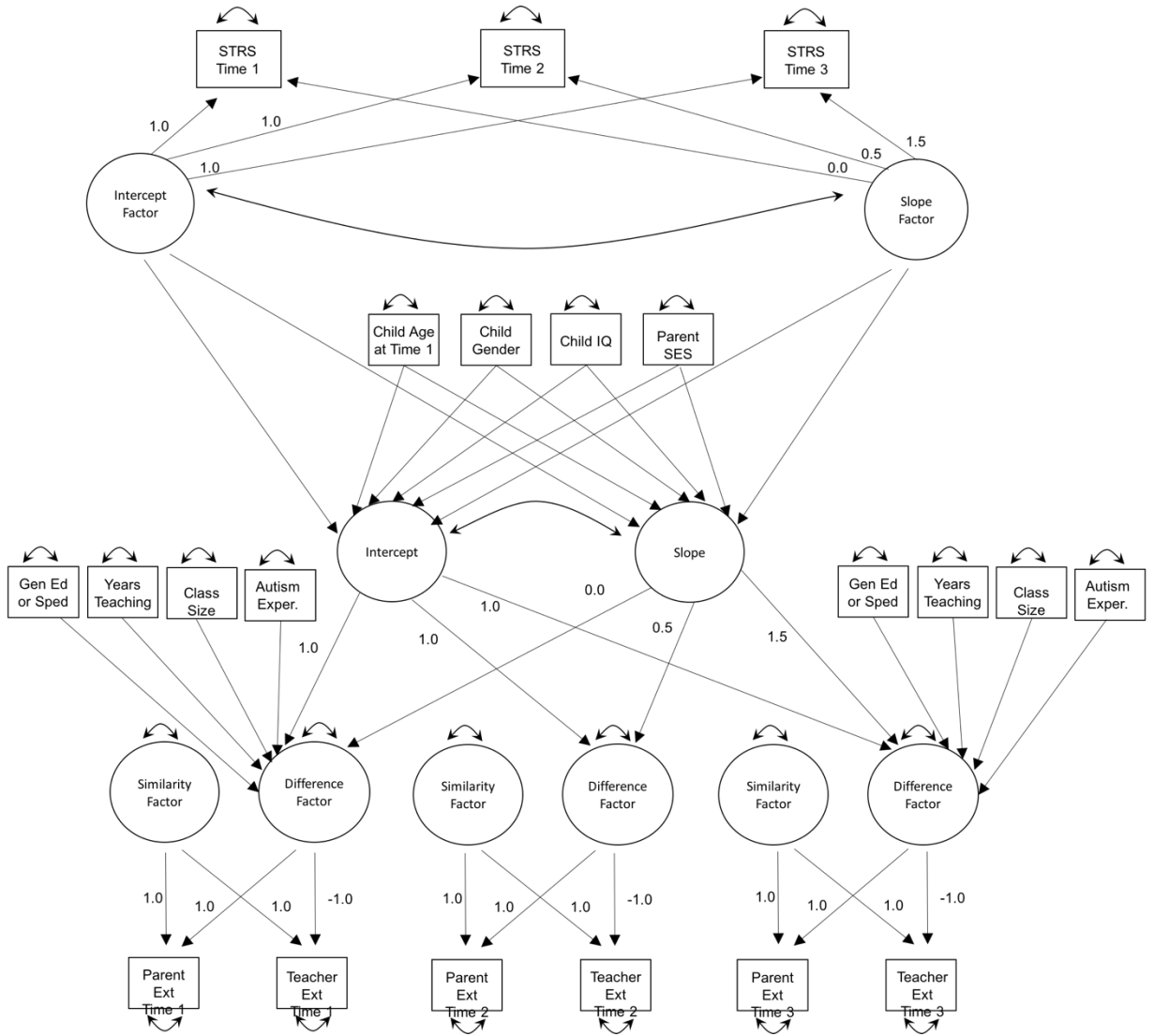discrepancies in parent and teacher-reported internalizing behavior across time
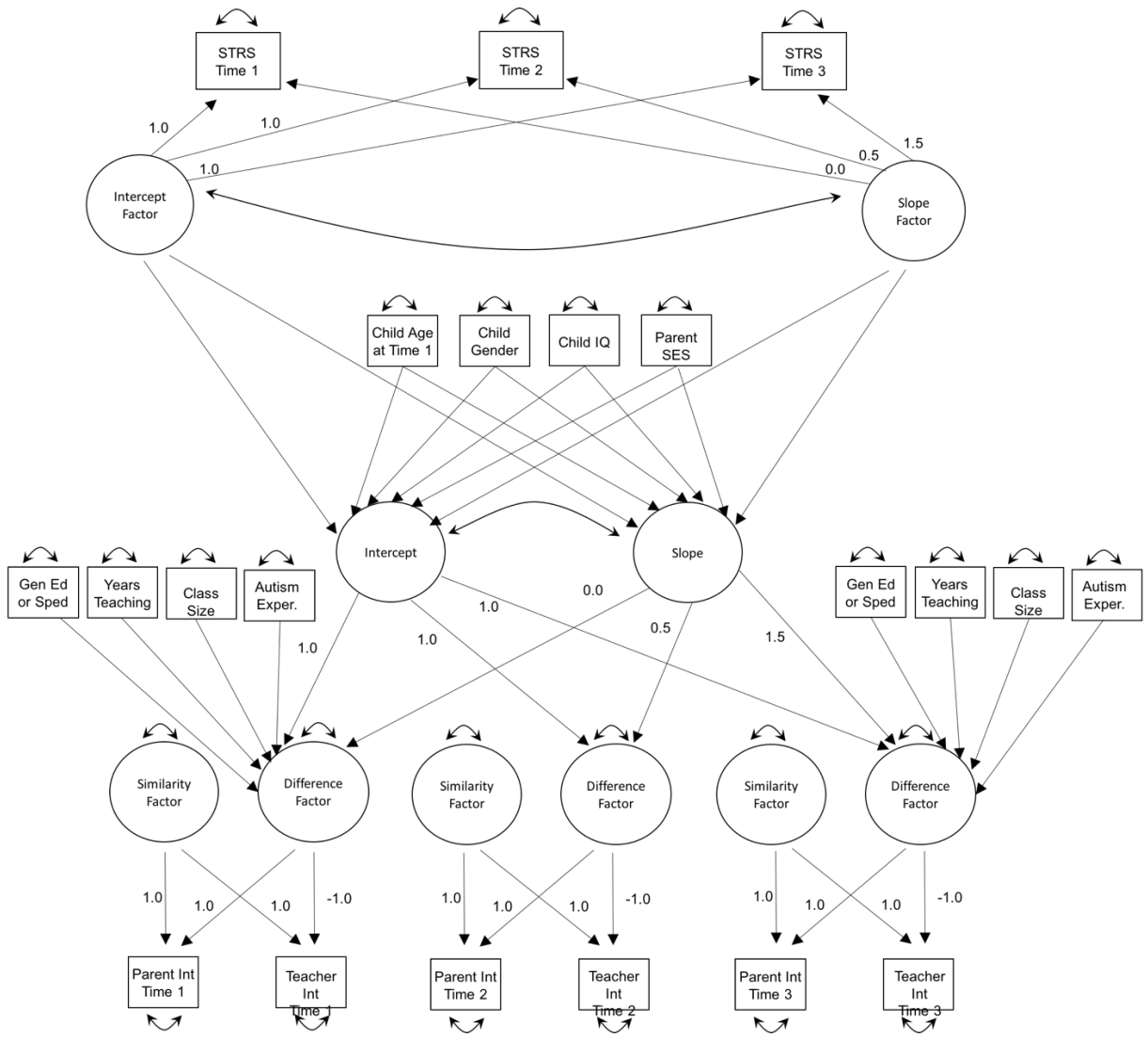
Figure 16.
Proposed model examining parent, child, and teacher characteristics that predict
discrepancies in parent and teacher-reported externalizing behavior across time
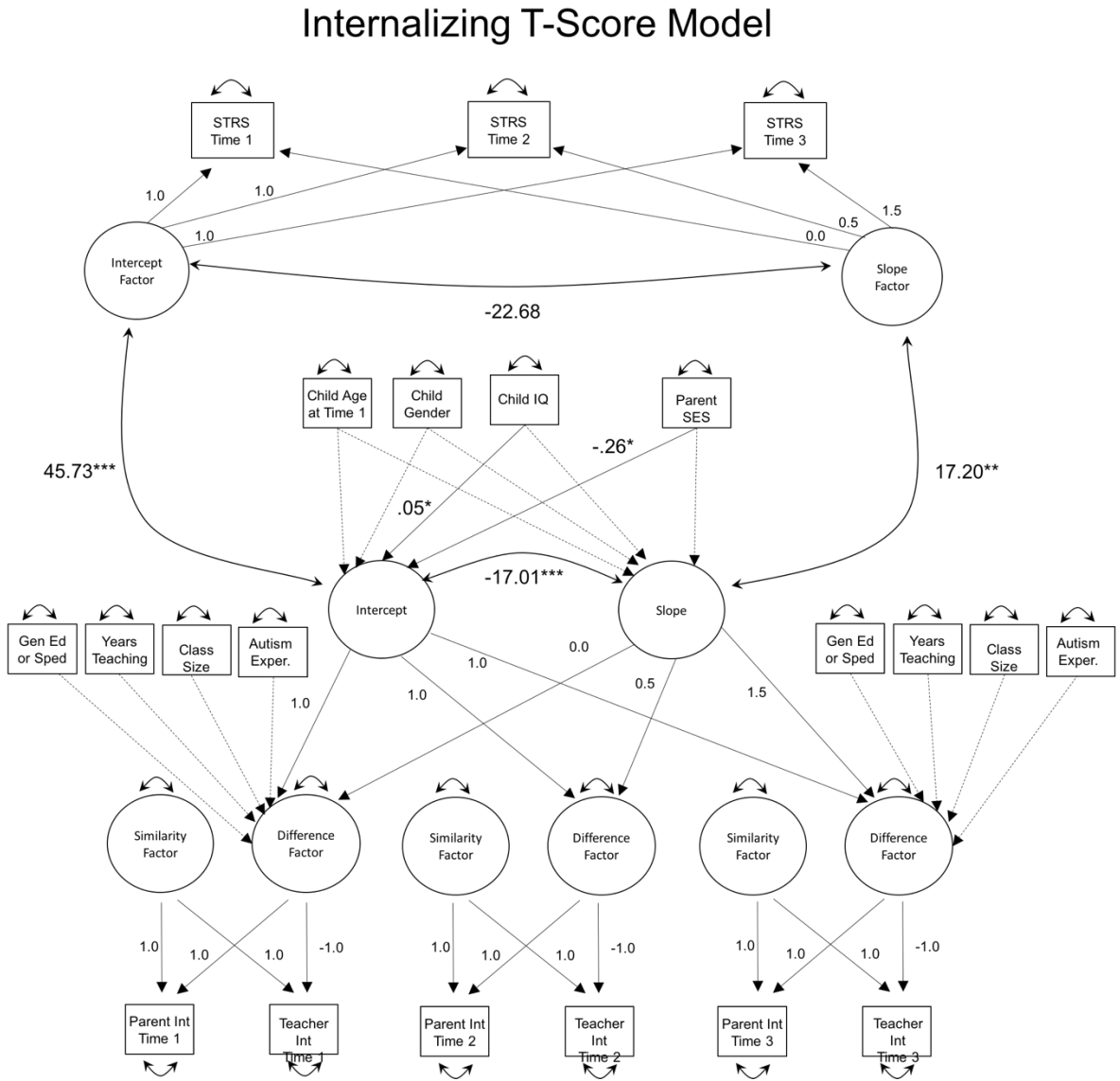
Figure 17.
Final Internalizing T-Score Model

Internalizing T-Score Model

Figure 18.
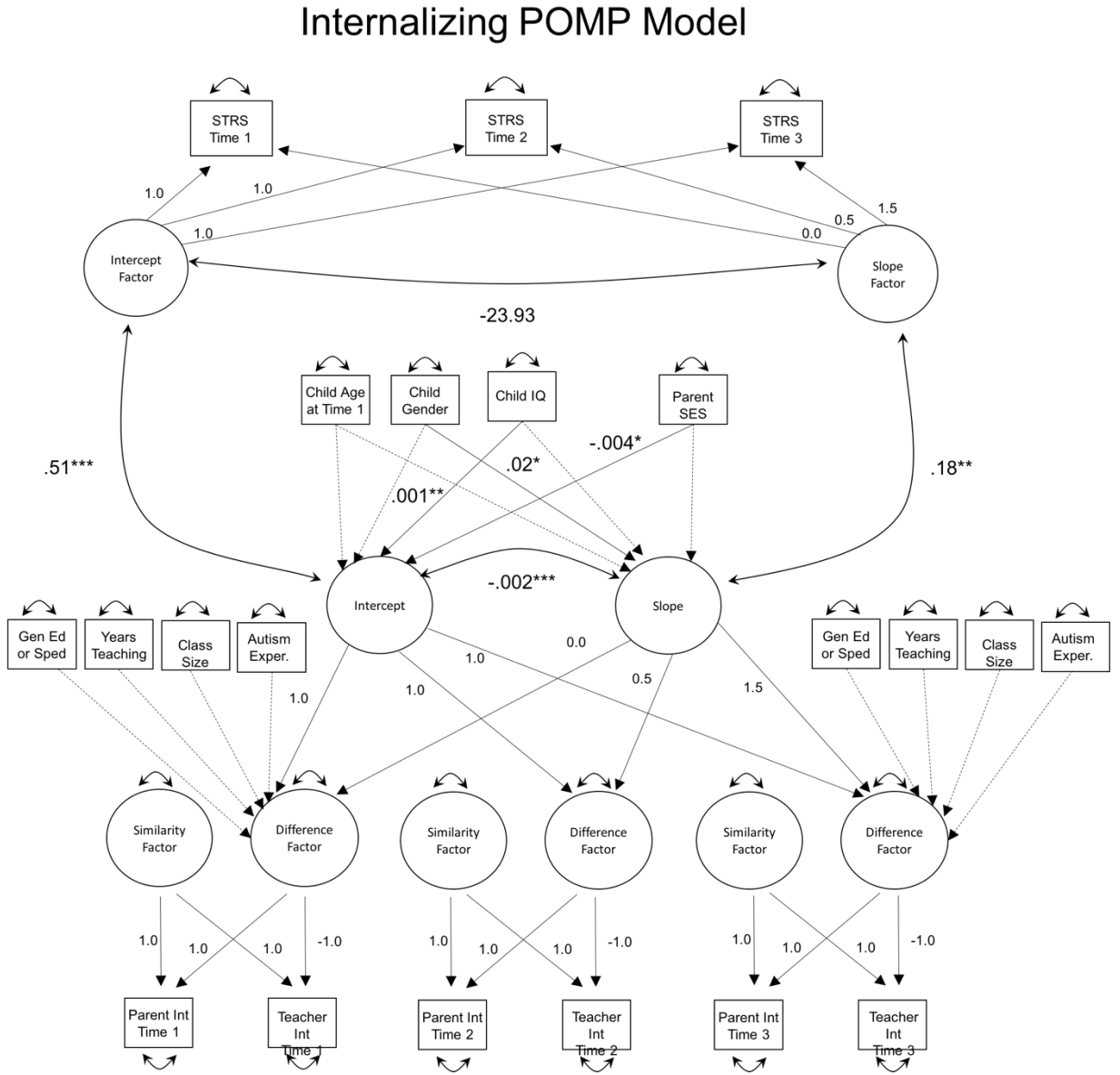Final Internalizing POMP Score Model
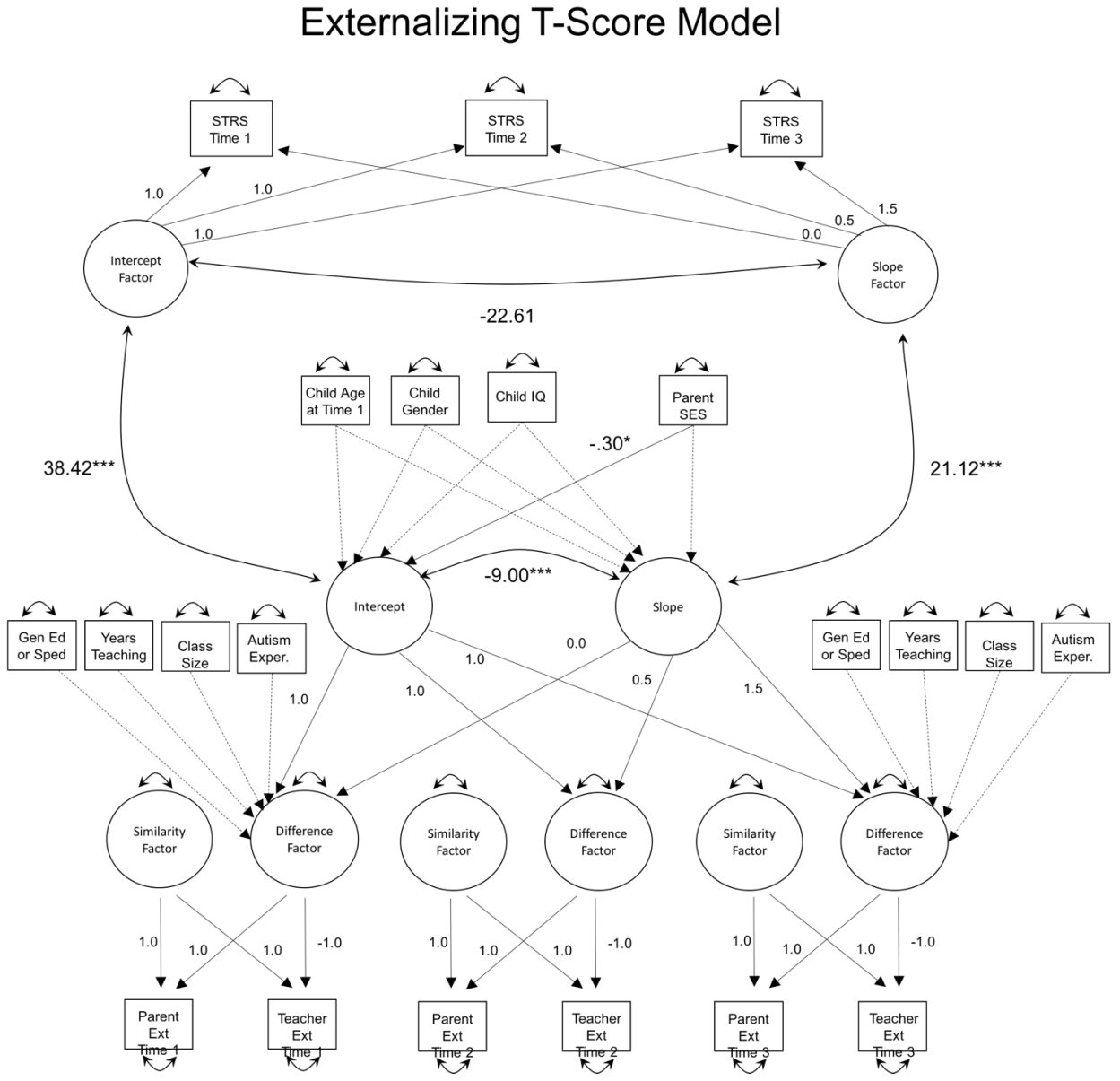
Figure 19.
Final Externalizing T-Score Model

Figure 20.
Final Externalizing POMP Score Model