



Comparing the Predictive Utility of Parent and Teacher Reports of Externalizing Behaviors on Concurrent Academic Achievement in Preschool-Aged Children

Christopher DeCamp¹ · Christopher J. Lonigan¹

Accepted: 17 October 2023 / Published online: 9 November 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Discrepancies between teacher and parent reports of children’s externalizing behaviors are well documented. However, less research has examined the associations these different ratings have with objective indicators of functioning in other domains. The goal of this study was to compare the strength of association of parent and teacher reports of externalizing behaviors with children’s early academic skills. The sample consisted of 695 children (376 boys, 318 girls, 1 unknown) who ranged between 48 months and 63 months of age (mean age=55.05; $SD=3.63$) at time of initial assessment. Children completed standardized measures of early academic skills; parents and teachers completed the Conners Rating Scale. Steiger’s Z tests were performed to compare the strength of associations between parent and teacher ratings on children’s early academic skills. Multi-level regressions examined the unique predictive variance each rater accounted for. Teacher ratings of inattentive and oppositional defiant behaviors had stronger associations with children’s early academic skills than did parent ratings for most measures of early academic skills, but there were no significant differences for ratings of hyperactive/impulsive behaviors. Multivariate analyses revealed that only teacher ratings of inattentive behaviors accounted for notable portions of unique variance in early academic skills. Children’s externalizing behaviors were related to their early academic skills. However, these results suggest that teachers contributed more unique variance, possibly due to their access to a normative reference group.

Keywords Externalizing Behaviors · Multi-modal Assessment · Behavior Report · Teacher Ratings · Parent Ratings · Early Academic Skills

The *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM; American Psychiatric Association, 2013) utilizes the consistency of behaviors associated with various disorders across settings for some diagnostic classifications. For Attention-Deficit/Hyperactivity Disorder (ADHD), symptoms of either inattention or hyperactivity/impulsivity must be present in at least two settings for the diagnostic criteria to be met; the number of symptoms present beyond the

diagnostic threshold indicates severity of the disorder. For Oppositional Defiant Disorder (ODD), symptoms of angry/irritable mood, argumentative/defiant behavior, and vindictiveness must be present in at least one setting; the number of settings in which these symptoms are present indicates severity of the disorder. Clinicians and researchers often rely on different reporters to assess the presence of symptoms across settings. For instance, children’s teachers report on child behaviors at school and children’s parents report on child behaviors at home. However, agreement between teachers and parents on reports of children’s externalizing behaviors is consistently reported to be low. Although there has been a substantial amount of research documenting these discrepancies (e.g., Achenbach et al., 1987; De Los Reyes et al., 2015), little prior work has explored the degree to which different raters’ reports of behavior correspond to objective indicators of functioning in other domains. Furthermore,

✉ Christopher DeCamp
decamp@psy.fsu.edu

Christopher J. Lonigan
lonigan@psy.fsu.edu

¹ Department of Psychology and the Florida Center for Reading Research, Florida State University, 1107 W. Call Street, Tallahassee, FL 32306-4301, USA

no prior work has explored whether these different raters' reports of behavior account for predominately unique or overlapping variance in such indicators. If raters have substantially different relations to an objective indicator that is negatively affected by ADHD or ODD symptoms, it raises questions about the broad use of multi-informant report in diagnostics. Exploring these questions may help clarify the importance of symptoms in different settings, clarify issues related to clinical utility of different raters, and provide a basis for selecting assessment and treatment targets in clinical settings.

Rater Discrepancies

A substantial body of evidence documents a lack of agreement between parent and teacher reports of externalizing behaviors (Antrop et al., 2002; De Los Reyes et al., 2013, 2015; Lane et al., 2013; Nelson & Harwood, 2011; Stone et al., 2013). A meta-analysis of 119 studies highlighted that parent and teacher reports had a small level of association ($r = .27$); however, the same meta-analysis indicated that reports of behavior from similar informants (e.g., two parents) had a relatively high level of association ($r = .60$; Achenbach et al., 1987). Corroborating the results of this older meta-analysis, results of more recent reviews also demonstrated that parent and teacher reports of various behaviors do not have high inter-rater reliability (De Los Reyes et al., 2013, 2015). For example, De Los Reyes et al. (2015) reported an average correlation of 0.28 between parent and teacher reports of externalizing behaviors across 162 studies published subsequent to the Achenbach et al. (1987) meta-analysis. Although some studies indicate that discrepancies are higher with younger children (e.g., Holmberg & Bolte, 2014), results of the De Los Reyes et al. (2015) meta-analysis and other studies (Narad et al., 2015) indicated discrepancies are consistent through adolescence. These modest correlations raise questions about the utility of either parent or teacher ratings.

De Los Reyes and colleagues (e.g., De Los Reyes & Epkins, 2023; De Los Reyes et al., 2013) noted that the traditional approach to discrepancies in reports of behavior is to focus on the degree of overlap. For example, variance common to two reports reflects the construct of interest and the non-overlapping variance reflects non-construct variability. De Los Reyes et al. (2013) introduced the Operations Triad Model (OTM) to shift the focus from “convergence as validity” toward a focus on understanding reasons for discrepancies. Within the OTM, there are three types of measurement operations: converging, diverging, and compensating. Converging operations reflect the traditional expectation that two measures of the same construct will agree, leading to

the same conclusion (e.g., psychopathology present, psychopathology not present). Diverging operations reflect instances when two measures do not agree, leading to different conclusions, but both reflect meaningful variation in the construct being assessed. Compensating operations reflect instances when two measures do not agree, leading to different conclusions, but the differences do not reflect meaningful variation in the construct being assessed because of methodological differences, error, or validity issues related to one or both measures. One goal of the OTM is to promote identification of situations in which discrepancies reflect meaningful variation and provide clinical utility.

The OTM provides a framework for considering when divergent ratings of behavior may be useful or when divergent ratings suggest a measurement or validity issue. It is possible that children's behavior is objectively different in different settings, and that teachers and parents accurately report the behavior they observe, resulting in discrepant ratings. This would be a case of diverging operations, and each rater's report would be expected to provide clinically useful information. It is also possible that there is variation in the validity of different raters' behavior ratings that yields discrepant ratings, which would indicate compensating operations. There are potential reasons for differences in teacher and parent ratings of objectively similar behaviors. First, because of their more extensive exposure to children, teachers have a better frame of reference for distinguishing typical and atypical behaviors for children of a certain age or developmental level than do parents; therefore, teachers' behavior ratings may be more accurate or more consistent from child to child. Second, behavioral expectations are likely to be stable across time and uniform across children in schools. Consequently, behavior that does not fall within expected behavioral norms is likely more salient, perhaps leading to more accurate or more consistent reporting. In contrast, behavioral expectations at home may be less stable across time (e.g., parents adapting expectations based on child behavior), less uniform across children (i.e., different parents have different expectations), or both. Thus, the identification of behavior that does not fall within expected behavioral norms may be idiosyncratic, leading to less accurate or less consistent reporting by parents. Finally, because of stable and uniform behavioral expectations in school, children face a greater demand for adaptation to those expectations than they do at home. In a child's home, behavioral expectations may be more likely to be adapted to the individual or, conversely, expectations may be held as rigid, which could lead to developmentally inappropriate expectations. Therefore, it is possible that in school settings, children unable to regulate behaviors to meet developmentally appropriate behavioral expectations are those who are accurately identified by teachers as exhibiting developmentally

atypical levels of externalizing behaviors, but children identified as not meeting behavioral expectations at home may not be those exhibiting developmentally atypical levels of externalizing behaviors.

Associations Between Externalizing Disorders and Other Domains

Externalizing diagnoses, such as ADHD and ODD, and the behaviors that define these disorders, are associated with impairments in other domains including academic performance (e.g., Allan et al., 2018; Evans et al., 2020; Gray et al., 2017; Kulkarni et al., 2020; Masten et al., 2005; Ogg et al., 2016; Owens et al., 2015; Owens et al., 2016; Tarver et al., 2014), social competence and peer relationships (e.g., Evans et al., 2020; Owens et al., 2015; Owens et al., 2016; Tarver et al., 2014), and other dimensions of self-regulation, like executive function (e.g., Romero-Lopez et al., 2017; Yang et al., 2022). In terms of the relation between externalizing behaviors and academic performance, both Arnold et al. (2020) and Frazier et al. (2007) conducted meta-analyses demonstrating that higher levels of ADHD symptoms were associated with lower levels of academic achievement, but these meta-analyses did not distinguish between inattentive and hyperactive/impulsive symptoms. Based on a qualitative synthesis of 41 studies, Gray et al. (2017) concluded that inattention in both clinical and non-clinical samples was significantly and negatively associated with academic outcomes. Symptoms of ODD are associated with poor academic performance, even when controlling for ADHD symptoms (Daley & Birchwood, 2010; Frazier et al., 2007), and have been associated with greater educational difficulties through high school (Burke et al., 2014).

Despite substantial discrepancies between teacher and parent report, studies tend to use either teacher or parent ratings, without acknowledging the reported discrepancies between different raters' reports or the possible differential associations with outcome domains (e.g., Burke et al., 2014; Fuchs et al., 2006; Swanson, 2011). Additionally, some studies do not differentiate between the two raters in analyses (e.g., Daucourt et al., 2020; Gray et al., 2017). For example, although the reviews by Gray (2017) and Daucourt et al. (2020) summarized studies that included both teacher and parent reports of children's ADHD-related behaviors, they did not examine teacher and parent report separately. However, examination of studies included in Gray and Daucourt et al. revealed that the correlation between teachers' ratings of ADHD-related behaviors and academic outcomes ranged from 0.15 to 0.60, with an average correlation of 0.47. In contrast, parents' ratings of ADHD-related behaviors and academic outcomes ranged from 0.17 to 0.37, with an average correlation of 0.25.

Despite discrepancies, separate raters' reports may still represent valid indicators of the presence or absence of various symptoms, as the different raters may be reporting on different aspects of the underlying psychopathology. If this were the case, it would be expected that reports of externalizing behaviors from different raters would have similar associations with other functional domains despite discrepancies between each other. However, if separate raters' ratings have differential relations to a functional outcome, such as academic achievement, that is negatively affected by the presence of externalizing behaviors and disorders, then raters may be supplying non-equivalent diagnostic information. Evidence suggests that requiring parents and teachers to agree on the presence of a behavior for an ADHD diagnosis reduces diagnostic accuracy (e.g., Martel et al., 2021). Although a comparison of the association between teacher and parent reports with outcomes in domains affected by a disorder is not a study of diagnostic accuracy, findings from such a study may have important implications for future clinical research, including the utility of different raters' reports for diagnoses.

Current Study

Whereas discrepancies between parent and teacher reports of externalizing behaviors have been well documented (e.g., Achenbach et al., 1987, De Los Reyes et al., 2015), much less research has examined the associations these different behavior reports have with academic achievement or other outcome domains. The research that has summarized the associations between behavior reports and academic achievement typically has not distinguished between parent or teacher reports (e.g., Gray et al., 2017; Daucourt et al., 2020). One way to address whether the discrepancies between parent and teacher reports of externalizing behaviors represent diverging operations or compensating operations is by comparing the associations of teacher and parent reports to another domain. Consequently, the primary goal of this study was to compare correlations between parent-rated externalizing behaviors and academic skills and teacher-rated externalizing behaviors and academic skills in a sample of preschool-age children. A secondary goal of this study was to examine the degree to which parent and teacher reports of externalizing behaviors yielded incremental predictive utility for academic achievement. Based on our analysis of data from studies included in the Gray et al. (2017) and Daucourt et al. (2020) reviews, it was hypothesized that teacher ratings of externalizing behaviors would have stronger predictive utility for academic achievement compared to parent reports of externalizing behavior. Based on the expectation that discrepancies between teacher and parent reports reflect diverging operations, it was hypothesized that

parent and teacher reports would have incremental predictive utility. Thus, when ratings from both parents and teachers were included in a predictive model, they would both account for unique as well as shared predictive variance.

Method

Participants

This study used an archival dataset from a large-scale study that involved the prediction and prevention of reading difficulties for preschool children who were at higher-than-average risk for such difficulties. Children were recruited from preschools in private centers and the Title I preschools of the local school district in north Florida in three successive years beginning in 2008. The sample used in this study included all children from the larger dataset who had parent ratings of behavior, teacher ratings of behavior, and academic outcome data. The sample included 695 children (376 boys, 318 girls, 1 unknown) who ranged in age from 48 months to 63 months (mean age = 55.05 months; $SD = 3.63$ months) at the time of their initial assessments. The sample was racially and ethnically diverse; 46.9% of the children were White, 43.2% were Black/African American, 3% were Hispanic, 2.7% were Asian, 0.3% were American Indian, 3.3% identified as more than one race, and the race of 0.4% of the sample was unknown. From the sample, 63.9% of participants' families earned \$50,000 or less, 14.8% of participants' families earned between \$50–\$75,000, 13.1% of participants' families earned between \$75–\$100,000, 6.4% of participants' families earned between \$100–\$150,000, and 1.7% of participants' families earned \$150,000 or more a year. The highest level of education for parents of participants in the sample was typically some college education (30.1%). However, 11.2% of participants had at least one parent who had earned an Associate's degree, 21.2% had earned a Bachelor's degree, 7.1% had earned a Master's degree, and 3.5% had earned a Doctoral degree. Some families had a high school diploma or GED as the highest level of education attained (16.2%); other families had some high school (8.6%) or no high school/less than 8th grade (0.6%).

Measures

Conners' Rating Scale for Preschool Children. Parents and teachers rated children's externalizing behaviors using either a hybrid version of the Conners' Rating Scale (CRS) that combined 44 items from multiple versions of the CRS (e.g., Conners, 1969, 1997; see Gerhardstein et al., 2003) or a 15-item version of the CRS developed by Purpura and Lonigan (2009) made up of a subset of the 44 items in the

hybrid version. Three dimensions represent items from both versions of the CRS: Inattention, Hyperactivity/Impulsivity, and Oppositional Defiant Behavior. This 15-item scale was created using item-response theory analysis to reduce the time needed for parents and teacher to complete the measures while retaining the scales' ability to discriminate children with different levels of behavioral problems along the three aforementioned factors. For both versions, parents and teachers rated behaviors on a 4-point Likert scale that ranged from zero, indicating that the behavior was not present at all, to three, indicating that the behavior was frequently present. For all participants, parent and teacher CRS scores were created by using the 15 items common to both the 44- and 15-item versions. Correlations between the 44- and 15-item versions are high ($r = .92$ for inattention subscale, 0.94 for hyperactive/impulsive subscale, and 0.96 for oppositional defiant subscale; Purpura & Lonigan, 2009). For both parents and teachers, the reliability of the CRS-15 ranged from acceptable to high for the Inattention (parent rating $\alpha = 0.81$; teacher rating $\alpha = 0.90$), Hyperactivity/Impulsivity ($\alpha = 0.81$; $\alpha = 0.75$), and Oppositional-Defiant ($\alpha = 0.78$; $\alpha = 0.88$) scales.

Preschool Comprehensive Test of Phonological and Print Processing. Participants were administered the *Preschool Comprehensive Test of Phonological and Print Processing* (Pre-CTOPPP; Lonigan et al., 2002), which measures three- to five-year-old children's phonological awareness, print knowledge, and oral language skills. The Pre-CTOPPP was the development version of the *Test of Preschool Early Literacy* (Lonigan et al., 2007). Children's phonological awareness skills were measured with two subtests: Blending and Elision. The Blending subtests consists of 21 items that required children to combine word sounds to form a word. The Elision subtest consists of 18 items that required children to say a word, then remove part of that word and say the new word. The Print Knowledge subtest consists of 36 items related to print concepts, letter name and letter sound recognition, and letter name and letter sound production. The Receptive Vocabulary subtest consists of 40 items for which children were required to point to a picture that represents a word spoken by the examiner. The Definitional Vocabulary subtest contains 40 items that require children to first provide the word that matches a picture and then provide information about the feature or function of the pictured item. All subtests include at least one practice item. Internal consistency for all Pre-CTOPPP subtests range from acceptable to high for three- to five-year-old children ($\alpha = 0.76 - 0.95$).

Test of Early Mathematics Ability—Third Edition. Children completed the *Test of Early Math Ability* (TEMA-3; Ginsburg & Baroody, 2003) to assess overall mathematics ability. The TEMA-3 consists of 72 items that are scored in a

binary, pass/fail fashion. The TEMA-3 produces an overall, math ability score but does not have sub-scales that assess specific dimensions of mathematical ability. The TEMA-3 has high internal consistency with three- ($\alpha = 0.92$) and four- ($\alpha = 0.93$) year-olds, as well as a two-week test-retest reliability of 0.82.

Child Math Assessment–Abbreviated. Children completed the *Child Math Assessment–Abbreviated* (CMA-A; Starkey et al., 2004) to assess their informal mathematical knowledge across a range of concepts. The CMA-A is composed of four sets of tasks that measure four dimensions of early mathematical knowledge. The first task has children solve simple addition and subtraction problems that involve a single set of objects that is initially invisible and then hidden from view. The second task has children construct a set of objects equal in number to a set that the children were shown. The third task asks children to recognize various shapes. The final task involves copying a repeating pattern using sets of objects that vary in color and identity from the objects in the model pattern. Each task is composed of multiple problems. The internal consistency of the CMA-A is acceptable for preschool children ($\alpha = 0.79$; Preschool Curriculum Evaluation Research Consortium, 2008).

Procedure

The current study was approved by the Florida State University Internal Review Board (HSC2010.0144). Parents of all participating children provided written informed consent/permission for their children’s participation, and all classroom teachers consented to participate. The data used for this study came from children’s assessment of academic skills during the fall of children’s preschool year. Children gave verbal assent prior to the beginning of testing. Parents and teachers completed the CRS in the fall of children’s preschool year (i.e., October through early-December), coincident with the assessment of children’s academic skills. Trained research assistants administered assessments to children in a quiet area of their preschools.

Research assistants were trained to criterion performance on the measures through didactic presentations, modeling, a performance assessment of test administration to an adult posing as a child, and live observations of assessment with feedback. Assessments were conducted over several 30- to 45-minute sessions, within a two- to three-week period.

Results

Some individuals ($n=38$) did not answer one item on the CRS. The items not answered differed across participants. Consequently, all scales were prorated, such that any child who had a score for at least four of the five items on a subscale still received a total subscale score that was utilized for analyses (i.e., available items within a scale were summed and then divided by the number of items summed). Correlations within and between, as well as descriptive statistics for, parent and teacher ratings of children’s externalizing behaviors are shown in Table 1. All behavior ratings were significantly correlated. Paired-samples *t*-tests were used to examine whether the mean scores from the three scales were different for parents and teachers. The results of these tests indicated that parents rated oppositional defiant, $t(685) = -6.55, p < .001$, and hyperactive/impulsive, $t(688) = -8.02, p < .001$, behaviors at significantly higher levels than did teachers, but inattentive behaviors, $t(691) = 0.38, p = .70$, were rated at similar levels by parents and teachers.

Zero-order correlations between parent and teacher ratings of behavior and all academic outcome variables are shown in Table 2. Teacher ratings of inattentive, hyperactive/impulsive, and oppositional defiant behaviors were all significantly associated with all academic outcomes. Parents’ ratings of inattentive and hyperactive/impulsive behaviors were significantly associated with all academic outcomes; parents’ ratings of oppositional defiant behaviors were only significantly associated with TEMA-3, print knowledge, and expressive vocabulary scores. For both parent and teacher ratings, correlations were nominally higher

Table 1 Correlations between and descriptive statistics of parent and teacher ratings of externalizing behaviors

Behavior Rating	Teacher Ratings			Parent Ratings		
	IA	H/I	ODB	IA	H/I	ODB
Teacher IA						
Teacher H/I	0.68***					
Teacher ODB	0.50***	0.61***				
Parent IA	0.39***	0.36***	0.26***			
Parent H/I	0.30***	0.30***	0.20***	0.72***		
Parent ODB	0.16***	0.20***	0.24***	0.59***	0.58***	
Mean	0.75	0.70***	0.61***	0.70	0.93***	0.78***
(SD)	(0.79)	(0.65)	(0.69)	(0.64)	(0.71)	(0.62)

Note. IA = Inattention; H/I = Hyperactivity/Impulsivity; ODB = Oppositional Defiant Behaviors; [†] $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$. Mean values with asterisks indicate a significant difference in mean ratings (i.e., parents rated the presence of significantly more H/I behaviors than teachers)

Table 2 Correlations of parent and teacher ratings of externalizing behavior domains with academic outcomes and comparisons of strength of parent and teacher correlations

Academic Measure	Teacher-Rating <i>r</i>	Parent-Rating <i>r</i>	Steiger's <i>Z</i> -value
Ratings of Inattentive Behavior			
TEMA-3	-0.44***	-0.28***	-4.12***
CMA-A	-0.45***	-0.22***	-5.93***
Blending	-0.21***	-0.16***	-1.08
Elision	-0.33***	-0.25***	-1.98*
Print Knowledge	-0.46***	-0.27***	-4.99***
Receptive Vocabulary	-0.34***	-0.18***	-4.03***
Expressive Vocabulary	-0.37***	-0.25***	-3.01**
Ratings of Hyperactive-Impulsive Behavior			
TEMA-3	-0.23***	-0.20***	-0.62
CMA-A	-0.21***	-0.14***	-1.49
Blending	-0.11**	-0.11**	-0.02
Elision	-0.18***	-0.22***	0.82
Print Knowledge	-0.22***	-0.20***	-0.39
Receptive Vocabulary	-0.20***	-0.13***	-1.66 ⁺
Expressive Vocabulary	-0.20***	-0.19***	-0.20
Ratings of Oppositional-Defiant Behavior			
TEMA-3	-0.22***	-0.08*	-3.17***
CMA-A	-0.21***	-0.04	-3.55***
Blending	-0.10**	-0.07 ⁺	-0.73
Elision	-0.19***	-0.06	-2.78**
Print Knowledge	-0.20***	-0.13**	-1.60
Receptive Vocabulary	-0.18***	-0.05	-2.86**
Expressive Vocabulary	-0.21***	-0.08*	-2.85**

Note. TEMA-3=Test of Early Math Abilities, Third Edition; CMA-A=Child Math Assessment, Abbreviated; Correlation between parent and teacher ratings: Inattention – $r = .40$; Hyperactive/Impulsive – $r = .30$; Oppositional Defiance – $r = .24$; ⁺ $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

for math, Print Knowledge, and language outcomes than they were for phonological awareness outcomes.

Comparing Predictive Utility of Parent and Teacher Reports

Steiger's *Z* tests (Steiger, 1980) were used to compare the strength of associations of parent and teacher ratings with academic outcomes, separated by behavior dimension and academic outcome (see Table 2). Results for inattentive behaviors revealed that teacher ratings had significantly stronger associations with all measures of academic outcomes, other than Blending, than did parent ratings. Results for hyperactive/impulsive behaviors revealed that there were no significant differences in associations between parent and teacher ratings for any academic outcome. Results for oppositional defiant behaviors produced results similar to those for inattentive behaviors such that teacher ratings had significantly stronger associations with academic

outcomes than did parent ratings for all academic outcomes except Blending and Print Knowledge.

Multivariate Prediction of Academic Outcomes

Multi-level regression analyses were conducted to determine if parent reports and teacher reports of children's externalizing behaviors accounted for portions of variance beyond each other, as well as child age at time of first assessment, child gender, child race, and highest education of parents. Children's preschools were included as a clustering variable to account for the nesting of these data. Because multi-level regression models do not include true R^2 values, pseudo- R^2 values were calculated using residuals from conditional and unconditional models to represent the total variance explained by full models (LaHuis et al., 2014). The pseudo- R^2 of control variables (i.e., age, gender, race, and highest education of parent; control block) and behavior variables (i.e., parent- and teacher-rated externalizing behaviors; behavior block) were calculated in a similar manner. Finally, unique predictive variance (sr^2) for all variables in a model was calculated using the change in pseudo- R^2 values when each variable was and was not included in a model (e.g., the pseudo- R^2 from the full model minus the pseudo- R^2 from a model without teacher-rated externalizing behaviors generates the sr^2 of teacher-rated externalizing behaviors). Predictive variance shared among variables can be computed by subtracting the sum of the sr^2 of the variables from one block (i.e., control block or behavior block) from the difference between the model pseudo- R^2 and the pseudo- R^2 of the other block.

Results of models that included children's early mathematics abilities are shown in Table 3. Nominally, ratings of inattentive behavior accounted for more variance in math abilities than did ratings of hyperactive/impulsive or oppositional defiant behaviors (see Table 3¹ for pseudo- R^2 values). Teacher-rated inattentive behaviors, hyperactive/impulsive behaviors, and oppositional defiant behaviors accounted for more unique variance than did parent-rated behaviors. For models including inattentive behaviors, the sr^2 value of teacher-rated behaviors was 0.10 for both the TEMA-3 and the CMA-A. In contrast, the sr^2 value of parent-rated behaviors was 0.01 for the TEMA-3 and 0.00 for the CMA-A. The variance unique to the behavior block was 0.15 and 0.13 for the TEMA-3 and CMA-A, respectively, indicating that parent and teacher ratings shared 3% of the variance accounted for in TEMA-3 and CMA-A scores. Finally, sr^2

¹ Because multi-level regression models were used, *b*-weights (i.e., unstandardized regression parameters) are reported in the tables. Interpretation of *b*-weights are in the units of measurement of the variables (e.g., a one-unit increase in teacher CRS inattention scores results in a 3.35 raw score decrease in CMA-A scores).

Table 3 Results of mixed-model regression using CRS scales from parents and teachers on math skills

Variables	TEMA-3				CMA-A			
	Pseudo- R^2	Pseudo- sr^2	b	(SE)	Pseudo- R^2	Pseudo- sr^2	b	(SE)
Ratings of Inattentive Behavior								
Control Block	0.08				0.09			
Chronological Age		0.04	0.31***	(0.05)		0.02	0.27***	(0.07)
Gender		0.00	0.04	(0.39)		0.00	0.79+	(0.48)
Parental Education		0.00	0.39**	(0.12)		0.01	0.65***	(0.14)
Race (White)		0.00	0.95	(0.68)		0.00	1.42+	(0.83)
Race (Black)		0.00	-0.75	(0.69)		0.00	-1.42+	(0.84)
Behavior Block	0.19				0.18			
Parent-Report		0.01	-1.05**	(0.33)		0.00	-0.56	(0.40)
Teacher-Report		0.10	-2.60***	(0.29)		0.10	-3.35***	(0.35)
All Variables	0.23				0.22			
Ratings of Hyperactive-Impulsive Behavior								
Control Block	0.08				0.07			
Chronological Age		0.06	0.39***	(0.06)		0.04	0.37***	(0.07)
Gender		0.00	0.37	(0.42)		0.00	1.30*	(0.52)
Parental Education		0.00	0.47***	(0.13)		0.02	0.78***	(0.15)
Race (White)		0.00	0.69	(0.74)		0.00	0.88	(0.90)
Race (Black)		0.00	-1.08	(0.75)		0.00	-1.98	(0.91)
Behavior Block	0.05				0.03			
Parent-Report		0.01	-0.91**	(0.30)		0.00	-0.45	(0.37)
Teacher-Report		0.02	-1.32***	(0.36)		0.01	-1.30**	(0.44)
All Variables	0.11				0.08			
Ratings of Oppositional-Defiant Behavior								
Control Block	0.09				0.09			
Chronological Age		0.07	0.41***	(0.06)		0.05	0.40***	(0.07)
Gender		0.00	0.84*	(0.41)		0.01	1.74***	(0.51)
Parental Education		0.01	0.56***	(0.13)		0.02	0.84***	(0.15)
Race (White)		0.00	0.76	(0.74)		0.00	1.02	(0.90)
Race (Black)		0.00	-0.72	(0.75)		0.00	-1.65+	(0.92)
Behavior Block	0.03				0.02			
Parent-Report		0.00	-0.09	(0.34)		0.00	0.26	(0.42)
Teacher-Report		0.02	-1.40***	(0.32)		0.02	-1.45***	(0.40)
All Variables	0.11				0.10			

Note. sr^2 = squared semi-partial correlation coefficient (unique variance); TEMA-3 = Test of Early Math Abilities, Third Edition; CMA-A = Child Math Assessment – Abbreviated; + $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

values for parent-rated and teacher-rated behaviors ranged between 0.00 and 0.02 for models including hyperactive/impulsive and oppositional defiant behaviors. Child race did not account for unique variance for TEMA-3 or CMA-A scores.

Similar multi-level regression analyses were conducted to determine if parent ratings and teacher ratings accounted for different amounts of variance in children’s early literacy skills (see Table 4). Nominally, ratings of inattentive behaviors accounted for more variance in early literacy skills than ratings of hyperactive/impulsive or oppositional defiant behaviors across all early literacy skills (i.e., Blending, Elision, and Print Knowledge subtests). The behavior block for the inattentive model accounted for 23% of the variance in Print Knowledge outcomes, 16% of which was unique variance. Thus, 4% of the unique variance captured by behavior ratings was shared between parent ratings and

teacher ratings. Additionally, the behavior block for inattentive behaviors accounted for 9% of the variance in Elision scores, 7% of which was unique. Teacher ratings contributed as much unique variance as variance they shared with parent ratings (3%), while parent ratings only accounted for 1% of the unique variance in that outcome. Teacher and parent ratings in all other models produced sr^2 values between 0.00 – 0.02. Consequently, these results indicated that teacher ratings of inattentive behavior displayed incremental utility on Print Knowledge scores. Child race did not account for unique variance for scores on any measure of early literacy skill.

Results for models predicting children’s language skills are shown in Table 5. Again, ratings of inattentive behaviors nominally accounted for more variance in language skills than did ratings of hyperactive/impulsive behaviors or oppositional defiant behaviors. Parent and teacher ratings

Table 4 Results of mixed-model regressions using CRS scales of parent and teacher report on early literacy skills

Variables	Blending			Elision			Print Knowledge		
	Pseudo-R ²	b	(SE)	Pseudo-R ²	b	(SE)	Pseudo-R ²	b	(SE)
Ratings of Inattentive Behavior									
Control Block	0.07			0.05			0.13		
Chronological Age	0.03	0.22 ^{***}	(0.04)		0.15 ^{***}	(0.04)		0.55 ^{***}	(0.09)
Gender	0.00	0.23	(0.32)		0.22	(0.27)		1.16 ⁺	(0.62)
Parental Education	0.01	0.30 ^{**}	(0.10)		0.17 ^{**}	(0.08)		1.04 ^{***}	(0.18)
Race (White)	0.00	1.28 [*]	(0.56)		1.34	(0.47)		3.01 ^{**}	(1.08)
Race (Black)	0.00	-0.62	(0.57)		-0.42	(0.47)		2.08 ⁺	(1.10)
Behavior Block	0.04			0.09			0.23		
Parent-Report	0.00	-0.55 ⁺	(0.27)		-0.90 ^{***}	(0.23)		-1.26 [*]	(0.52)
Teacher-Report	0.02	-0.72 ^{**}	(0.24)		-1.01 ^{***}	(0.20)		-4.51 ^{***}	(0.45)
All Variables	0.10			0.12			0.29		
Ratings of Hyperactive-Impulsive Behavior									
Control Block	0.07			0.05			0.12		
Chronological Age	0.04	0.24 ^{***}	(0.04)		0.18 ^{***}	(0.04)		0.69 ^{***}	(0.09)
Gender	0.00	0.35	(0.33)		0.34	(0.28)		1.85 ^{**}	(0.69)
Parental Education	0.01	0.32 ^{**}	(0.10)		0.18 [*]	(0.08)		1.23 ^{***}	(0.20)
Race (White)	0.00	1.19 [*]	(0.57)		1.19 [*]	(0.48)		2.71 [*]	(1.18)
Race (Black)	0.00	-0.72	(0.58)		-0.61	(0.49)		1.55	(1.20)
Behavior Block	0.02			0.04			0.05		
Parent-Report	0.00	-0.41 ⁺	(0.23)		-0.91 ^{***}	(0.20)		-1.37 ^{**}	(0.49)
Teacher-Report	0.00	-0.38	(0.28)		-0.41 ⁺	(0.24)		-1.89 ^{**}	(0.57)
All Variables	0.08			0.08			0.15		
Ratings of Oppositional-Defiant Behavior									
Control Block	0.06			0.05			0.13		
Chronological Age	0.04	0.25 ^{***}	(0.04)		0.18 ^{***}	(0.04)		0.72 ^{***}	(0.09)
Gender	0.00	0.54 ⁺	(0.32)		0.64 [*]	(0.27)		2.52 ^{***}	(0.67)
Parental Education	0.01	0.34 ^{**}	(0.10)		0.24 ^{**}	(0.08)		1.32 ^{***}	(0.20)
Race (White)	0.00	1.20 [*]	(0.57)		1.22 [*]	(0.49)		2.69 [*]	(1.18)
Race (Black)	0.00	-0.67	(0.58)		-0.40	(0.50)		1.86	(1.20)
Behavior Block	0.00			0.01			0.03		
Parent-Report	0.00	-0.28	(0.25)		-0.10	(0.23)		-0.72	(0.56)
Teacher-Report	0.00	-0.20	(0.27)		-0.58 ^{**}	(0.21)		-2.02 ^{***}	(0.52)
All Variables	0.07			0.05			0.15		

Note. *svr*² = squared semi-partial correlation coefficient (unique variance); ⁺*p* < .10; ^{*}*p* < .05; ^{**}*p* < .01; ^{***}*p* < .001.

of hyperactive/impulsive behavior and oppositional defiant behavior yielded sr^2 values between 0.00 and 0.01. Teacher-rated inattention contributed unique variance to both Receptive Vocabulary and Expressive Vocabulary scores, whereas parent-rated inattention produced negligible sr^2 values. The inattentive-behavior block accounted for 9% and 12% of variance in Receptive Vocabulary and Expressive Vocabulary subtests, respectively, and 7% and 10% of that variance was unique variance, respectively. Similar to other academic outcomes, all the variance that parent ratings accounted for in Receptive Vocabulary and all but 1% of the variance in Expressive Vocabulary was shared with teacher ratings and the control block. Conversely, teacher-rated inattention accounted for 6% and 5% of the unique variance in Receptive and Expressive Vocabulary subtests, respectively. Consequently, only teacher ratings of inattention displayed incremental predictive utility for Receptive

and Expressive Vocabulary scores. Child race accounted for marginal amounts of unique variance in language skills; the White racial category contributed 1% unique variance to the receptive and expressive vocabulary in inattentive models as well as for receptive vocabulary in hyperactive/impulsive and oppositional defiant models.

Discussion

Although multi-informant assessment of children’s externalizing behavior problems is considered critical for the assessment of some mental health disorders (e.g., ADHD, ODD; American Psychiatric Association, 2013), parent and teacher ratings of externalizing behavior have consistently low levels of agreement (Achenbach et al., 1987; De Los Reyes et al., 2015). Such findings could indicate that both

Table 5 Results of mixed-model regression using CRS scales of parent and teacher report on early language skills

Variables	Receptive Vocabulary				Expressive Vocabulary			
	Pseudo- R^2	Pseudo- sr^2	b	(SE)	Pseudo- R^2	Pseudo- sr^2	b	(SE)
Ratings of Inattentive Behavior								
Control Block	0.06				0.07			
Chronological Age		0.02	0.23***	(0.05)		0.02	0.59***	(0.12)
Gender		0.00	0.24	(0.36)		0.00	0.51	(0.87)
Parental Education		0.01	0.46***	(0.11)		0.01	1.11***	(0.26)
Race (White)		0.01	2.08**	(0.64)		0.01	5.47***	(1.51)
Race (Black)		0.00	0.20	(0.64)		0.00	-0.09	(1.53)
Behavior Block	0.09				0.12			
Parent-Report		0.00	-0.32	(0.31)		0.01	-2.47**	(0.73)
Teacher-Report		0.06	-1.77***	(0.27)		0.05	-4.22***	(0.64)
All Variables	0.13				0.17			
Ratings of Hyperactive-Impulsive Behavior								
Control Block	0.06				0.07			
Chronological Age		0.03	0.27***	(0.05)		0.04	0.71***	(0.12)
Gender		0.00	0.35	(0.38)		0.00	1.03	(0.92)
Parental Education		0.02	0.52***	(0.11)		0.01	1.23***	(0.27)
Race (White)		0.01	1.98**	(0.66)		0.00	4.92**	(1.59)
Race (Black)		0.00	0.06	(0.67)		0.00	-0.65	(1.62)
Behavior Block	0.02				0.04			
Parent-Report		0.00	-0.40	(0.27)		0.01	-2.07**	(0.65)
Teacher-Report		0.01	-1.09***	(0.32)		0.01	-2.14**	(0.78)
All Variables	0.07				0.09			
Ratings of Oppositional-Defiant Behavior								
Control Block	0.06				0.07			
Chronological Age		0.03	0.29***	(0.05)		0.04	0.75***	(0.12)
Gender		0.00	0.73*	(0.37)		0.00	1.94*	(0.90)
Parental Education		0.02	0.57***	(0.11)		0.02	1.40***	(0.27)
Race (White)		0.01	1.93**	(0.66)		0.00	4.90**	(1.60)
Race (Black)		0.00	0.11	(0.67)		0.00	-0.18	(1.63)
Behavior Block	0.01				0.02			
Parent-Report		0.00	0.09	(0.31)		0.00	-0.27	(0.75)
Teacher-Report		0.01	-0.91**	(0.29)		0.01	-2.55***	(0.70)
All Variables	0.07				0.08			

Note. sr^2 = squared semi-partial correlation coefficient (unique variance); * $p < .10$; ** $p < .05$; *** $p < .01$; **** $p < .001$

parents and teachers provide unique and useful information about children's behavior. Alternatively, differences between raters could indicate issues related to the validity or utility of one or the other rater. Studies in which behavior ratings are used to predict other outcomes often treat parent and teacher ratings as interchangeable (e.g., Daucourt et al., 2020; Gray et al., 2017). In this study, which directly compared the predictive utility of parent and teacher ratings for academic outcomes, teacher ratings were generally more strongly associated with academic outcomes than were parent ratings, and only teacher ratings provided consistent and non-trivial unique predictive information. These results indicate that there may not be uniform utility in multi-informant assessment.

Although teacher ratings of externalizing behaviors were more strongly associated with children's academic outcomes than were parent ratings of the same behaviors, this finding was not consistent across domains of externalizing behavior or across domains of academic outcomes. For both teachers and parents, ratings of inattentive behaviors were more strongly associated with academic outcomes than were ratings of hyperactive/impulsive behaviors and oppositional defiant behaviors. Results of Steiger's *Z* tests provided partial support for the primary hypothesis that teacher ratings would have stronger associations with academic outcomes than would parent ratings. Although teacher ratings of hyperactive/impulsive behavior had nominally larger correlations than did parent ratings, there were no significant differences in correlations for any of the seven academic outcome measures. It is unclear why there were significant differences in associations for academic outcomes between teacher and parent ratings of oppositional defiant behaviors but not hyperactive/impulsive behaviors. It is, however, notable that teacher-parent agreement on oppositional defiant behaviors was the lowest among the externalizing behavior domains examined. For both parent and teacher ratings, correlations were nominally higher for math, Print Knowledge, and language outcomes than they were for phonological awareness outcomes.

Consistent with results from univariate analyses and in partial support of the second hypothesis, teacher ratings contributed more unique variance to academic outcomes than did parent ratings in multivariate analyses. Multi-level regression analyses revealed that there was partial overlap between parent and teacher ratings, as indicated by non-zero amounts of shared predictive variance for teacher and parent ratings. Further mirroring results of univariate analyses, teacher and parent ratings of inattentive behaviors accounted for more variance in children's academic outcomes than did ratings of hyperactive/impulsive or oppositional defiant behaviors. Control variables accounted for larger portions of variance than did behavior ratings for models examining

ratings of hyperactive/impulsive and oppositional defiant behaviors. Parent ratings only contributed more unique variance than did teacher ratings of the same behaviors on one occasion, and parent ratings often contributed no unique variance when considering teacher ratings and the control variables. Conversely, teacher-rated inattention contributed more unique variance than any other variable for six out of the seven measures of academic outcomes examined. Consequently, our findings only support teacher-rated inattention as consistently displaying incremental predictive utility. However, the total amount of variance explained was modest, and more variability in academic outcomes was unexplained than accounted for by the combination of parent and teacher ratings and the control variables.

Results of multivariate analyses were in-line with prior literature. For example, Gray and colleagues (2017) reported that ratings of inattention accounted for between 5% and 16% of the variance in standardized measures of academic outcomes, and, in the current study, the behavior block for inattentive models accounted for 4–23% of the variance in standardized academic outcomes. Some existing research supports the notion that inattentive behaviors, more so than hyperactive/impulsive behaviors or oppositional defiant behaviors, form a distinct risk factor for poor academic outcomes in children (e.g., Massetti et al., 2008). Notably, there were differences in strengths of associations depending on the academic outcome. This could be due to inattentive behaviors playing a larger role in the development of some academic skills than others. Thus, the present findings corroborate some existing research in supporting inattentive behaviors having stronger associations with academic outcomes than other domains of externalizing behaviors and are thus a more salient risk factor for worse academic outcomes than other domains of externalizing behaviors. Future research should further explore the differential relations of inattentive behaviors with specific academic skills.

Compensating or Divergent Operations?

The modest associations between parent and teacher reports in this study were consistent with prior studies (e.g., Achenbach et al., 1987; De Los Reyes et al., 2015) and may indicate validity issues. Despite parent and teacher ratings of child behaviors having differential associations with an objective outcome measure (i.e., academic outcomes), both parents' and teachers' ratings accounted for unique variance in some of those outcomes (i.e., 33% and 81% of the 21 behavior-domain-by-academic-outcome models for parent ratings and teacher ratings accounted for non-zero portions of variance, respectively). However, some of the variance accounted for was shared between parent and teacher ratings. Thus, there were not many cases where both raters

were providing novel predictive information. Given the substantive differences in unique variance accounted for by teacher ratings, the results of this study suggest that there may be limited value in including parent ratings in models to understand children's academic outcomes. Teachers' ratings provided novel information more often than did parents' ratings, and, in cases where there was shared predictive variance, it was generally less than, or equal to, the unique variance provided by parent ratings. Even for inattentive models, in which behavior ratings captured more variance compared to the models that included other domains of externalizing behaviors, the amount of variance shared between parent and teacher ratings always exceeded the amount of variance parent ratings uniquely captured.

This study focused solely on academic outcomes as an objective indicator of functioning, which leaves open the possibility that the obtained pattern of results is relevant for only this domain of functioning. That is, because teachers rate children's behaviors in the context of learning (i.e., school), teacher ratings are better indicators of academic success, whereas parents rate children's behaviors outside of school and thus may be better indicators of non-academic functioning, such as executive function or social skills. Alternatively, teachers' ratings may reflect implicit explanations for why a child is doing worse academically than other children (i.e., "this student does not pay attention" "this student is impulsive"); in this case, the academic performance, and not the child's behavior, drives the teacher's rating. However, teachers' ratings of behavior as a proxy for children's academic competencies seems an unlikely explanation in this study. First, ratings were collected in the fall of the school year, which would limit (but not eliminate) the amount of time that behaviors could interfere with learning. Second, teachers' ratings were more predictive of academic outcomes than were parent ratings across most outcome measures, including measures of language, and preschool exposure seems to have a larger effect on non-language skills than on language skills (e.g., Ansari et al., 2020). Finally, preschools do not have a strong emphasis on direct instruction as is expected in later grades, and the quality of language instruction is often judged to be low (e.g., Justice et al., 2008).

As noted previously, the OTM was intended to shift focus away from convergence as validity toward understanding reasons for discrepancies and identifying when discrepant results provide information with clinical utility. Compensating operations describe situations in which measures do not agree, leading to different conclusions, but the differences between measures do not reflect meaningful variation in the construct of interest. Compensating operations may be due to methodological differences, error, or validity issues related to one or both measures. In this study, teacher

and parent ratings had similar reliabilities, indicating that measurement error was unlikely to be a significant factor in discrepancies. Jungersen and Lonigan (2021) evaluated the measurement invariance of the parent and teacher report on the CRS and found that teachers and parents were reporting on largely the same constructs (i.e., partial metric invariance), and that teachers and parents were reporting differences in levels of the underlying constructs in similar ways (i.e., partial scalar invariance). These results suggest that discrepancies between raters are not the result of teachers and parents interpreting the externalizing behavior constructs differently.

Within the OTM, diverging operations describe situations in which measures do not agree, leading to different conclusions, but both measures reflect meaningful variation in the construct being assessed. Diverging operations can be due to factors such as unique perspectives that may result from observing events or individuals in different contexts or having context-related understandings of behavior. Although teachers' ratings of externalizing behaviors being a stronger correlate of preschool children's academic skills and having consistent unique predictive utility above parent ratings could be a function of such context-related factors or teachers' context-related understanding of behavior, such an explanation seems less likely given findings from studies examining functional outcomes other than academic skills.

Studies that directly compare the utility of teacher and parent ratings of children's behavior for the prediction of some independent, objective outcome are relatively rare; however, available evidence indicates a general pattern of stronger associations between teachers' ratings and measures of other outcomes than the associations between parents' ratings and that outcome. For instance, in terms of non-rating-based measures of inattention, hyperactivity/impulsivity, or both, Wang et al. (2015) reported that teacher ratings of inattention had the strongest correlations with scores on the *Test of Variables of Attention* compared to parent or clinician ratings of inattention. Similarly, based on a relatively large sample of Korean third- and fourth-grade children, Cho et al. (2011) reported a correlation between teacher-rated inattention and omission errors on a continuous performance task (CPT) that was significantly higher than the correlation between parent-rated inattention and omission errors, and the correlation between teacher-rated hyperactivity/impulsivity and commission errors on a CPT was significantly stronger than the correlation for parent-rated hyperactivity/impulsivity and commission errors.

Studies have reported a similar pattern of results for measures of executive function. Backer-Grondahl et al. (2019) noted that teacher-reported externalizing behaviors were associated with their measure of effortful control (EC;

similar to inhibitory control) but parent-reported externalizing behaviors were not related to their measure of EC. Data from the study by Cho et al. (2011) indicated that teacher ratings of inattention were more strongly associated with scores on a Stroop task and interference on the Children's Color Trails Test than were parent ratings of inattention. Jungersen and Lonigan (2020) reported that first- and second-grade teachers' ratings of inattention and hyperactivity/impulsivity on two different measures of externalizing behaviors were more strongly correlated with a latent variable representing six executive function tasks than were parents' ratings on the same measures.

The present findings coupled with existing research suggest that the context in which the child is observed is not the primary cause of observed differential associations. If children are behaving differently in different contexts, those behavioral discrepancies would cause mean rating differences, which would then lead to significant differences in associations with outcomes. However, mean ratings of inattentive behaviors, the domain for which differences in associations with academic outcomes were the largest, were not significantly different between teachers and parents, but parents were rating significantly more hyperactive/impulsive behaviors than teachers, the domain for which there were no significant differences in associations with academic outcomes. Thus, it becomes more likely that parents do not have a strong understanding of when various behaviors are problematic or what behaviors may be age appropriate because parents observe fewer children and likely only have many observations of their own children. This may lead to developmentally inappropriate expectations of a child that cause parents to erroneously identify developmentally typical levels of externalizing behaviors as problematic. Conversely, teachers have classrooms that consist of children of varying temperaments and achievement levels. Therefore, teachers may understand what behaviors are age-appropriate and when the presence of specific behaviors can have stronger impediments on relevant functional outcomes. Such potential differences in parents' and teachers' experiences could be a reason for the observed differences in parents' and teachers' behavior ratings, which could represent a diverging operation. However, even if raters differ for meaningful reasons, such as understanding when behaviors are more likely to cause an impediment, each rater would be expected to provide unique information. Considering that parent ratings accounted for at most 2% of the unique variance of a given outcome, parents' ratings did not contribute much beyond teachers' ratings. Because parent ratings did not contribute non-trivial amounts of unique variance beyond teacher ratings and there is no evidence to support measurement error, rater discrepancies likely reflect some other form of compensating operations.

Limitations and Future Directions

This study had numerous strengths, including a large and racially diverse sample and multiple standardized measures of children's academic skills. Despite these strengths, some limitations were present. This study only examined concurrent associations between parent and teacher ratings and children's academic abilities. Future research should examine predictions of longitudinal outcomes. Additionally, this study used only academic achievement data as an outcome measure. Future studies should examine the associations of teacher and parent ratings with multiple functional outcomes to assess whether raters account for largely overlapping, or largely unique, portions of variance. Moreover, comparing parent and teacher ratings' associations with direct behavioral observations, performed by a researcher who was trained to some criterion, as well as assessing behavioral ratings' associations with ratings of academic competency, could further the understanding of rater discrepancies.

Conclusions

The results of this study, when put into context with existing research, suggest that the information provided by different informants does not have universal utility across outcomes. Clinicians utilize multi-informant reports for various diagnoses that require consistency of symptoms across settings (e.g., ADHD, ODD), but there is no current, evidence-based, gold-standard practice by which to utilize multi-informant reports. Because diagnoses require consistency of symptoms across settings, one might think that diagnosis of such a disorder requires both teachers and parents to endorse the same behavior. However, such a designation is not required; clinicians and researchers employ varying strategies to combine the information to incorporate reports from multiple informants. Additionally, some research suggests that requiring a parent and teacher to agree on the presence of a given behavior results in worse negative predictive value, sensitivity, specificity, and similar positive predictive value (100 for both, 99.1; for either) compared to considering either parent or teacher endorsing a given behavior for the diagnosis of ADHD, when using a diagnosis made by a clinician as the standard to which these methods were compared (Martel et al., 2021).

Although many studies have compared parent and teacher reports of children's externalizing behaviors to each other, as well as assessed the predictive utility of either parent or teacher ratings on some objective domain of functioning, no studies have directly compared the predictive utility of these different raters on academic achievement to assess each rater's unique contributions. Results of this study

provided partial support for the primary hypothesis; teacher ratings of inattentive and oppositional defiant behaviors had consistently stronger correlations with academic outcomes than did parent ratings. The results of the current study provided some support for the second hypothesis; teacher ratings more consistently contributed unique predictive variance to academic outcomes than did parent ratings. This study provided evidence that teacher ratings generally captured more unique variance in academic outcomes than the variance they shared with parent ratings or the variance that parent ratings uniquely captured. When considering existing research that highlights teacher ratings are also a stronger correlate of executive function measures compared to parent ratings (Cho et al., 2011; Backer-Grondahl et al., 2019), these results are not likely to be explained by context driving the observed differential associations. Thus, these findings open the possibility that teachers' ratings provide more useful information than parent ratings due to teachers' ability to identify normative behaviors. Consistent with OTM, developing a better understanding of the outcomes for which discrepancies reflect meaningful variation and provide clinical utility can improve diagnostic accuracy.

Funding Portions of this research and report were supported by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (HD052120) and the Institute of Education Science (R305B200020). The views expressed herein are those of the authors and have not been reviewed or approved by the granting agencies.

Conflict of Interest None.

References

- 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*(2), 213–232.
- Allan, D. M., Allan, N. P., Lonigan, C. J., Hume, L. E., Farrington, A. L., & Vinco, M. H. (2018). The influences of multiple informants' ratings of inattention on preschoolers' emergent literacy skills growth. *Learning and Individual Differences*, *65*, 90–99. <https://doi.org/10.1016/j.lindif.2018.05.014>.
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>.
- Ansari, A., Pianta, R. C., Whittaker, J. V., Vitiello, V. E., & Ruzek, E. A. (2020). Persistence and convergence: The end of kindergarten outcomes of pre-k graduates and their nonattending peers. *Developmental Psychology*, *56*(1), 2027–2039. <https://doi.org/10.1037/dev0001115>.
- Antrop, I., Roeyers, H., Oosterlaan, J., & Van Oost, P. (2002). Agreement between parent and teacher ratings of disruptive behavior disorders in children with clinically diagnosed ADHD. *Journal of Psychopathology and Behavioral Assessment*, *24*(1), 67–73.
- Arnold, L. E., Hodgkins, P., Kahle, J., Madhoo, M., & Kewley, G. (2020). Long-term outcomes of ADHD: Academic achievement and performance. *Journal of Attention Disorders*, *1*, 1–13. <https://doi.org/10.1177/1087054714566076>.
- Backer-Grondahl, A., Naere, A., & Idsoe, T. (2019). Hot and cool self-regulation, academic competence, and maladjustment: Mediating and differential relations. *Child Development*, *90*(6), 2171–2188. <https://doi.org/10.1111/cdev.13104>.
- Burke, J. D., Rowe, R., & Boylan, K. (2014). Functional outcomes of child and adolescent oppositional defiant disorder symptoms in young adult men. *Journal of Child Psychology and Psychiatry*, *55*(3), 264–272. <https://doi.org/10.1111/jcpp.12150>.
- Cho, S. C., Kim, H. W., Kim, B. N., Shin, M. S., Yoo, H. J., Kim, J. W., Bhang, S. Y., & Cho, I. H. (2011). Are teacher ratings and parent ratings differentially associated with children's intelligence and cognitive performance? *Psychiatry Investigation*, *8*(1), 15–21. <https://doi.org/10.4306/pi.2011.8.1.15>.
- Conners, C. K. (1969). A teacher rating scale for use in drug studies with children. *American Journal of Psychiatry*, *126*(6), 884–888.
- Conners, C. K., Wells, K. C., Parker, J. D., Sitarenios, G., Diamond, J. M., & Powell, J. W. (1997). A new self-report scale for assessment of adolescent psychopathology: Factor structure, reliability, validity, and diagnostic sensitivity. *Journal of Abnormal Child Psychology*, *25*(6), 487–497.
- Daucourt, M. C., Erbeli, F., Little, C. W., Haughbrook, R., & Hart, S. A. (2020). A meta-analytical review of the genetic and environmental correlations between reading and attention-deficit/hyperactivity disorder symptoms and reading and math. *Scientific Studies of Reading*, *24*(1), 23–56. <https://doi.org/10.1080/10888428.2019.1631827>.
- De Los Reyes, A., & Epkins, C. C. (2023). Introduction to the special issue. A dozen years of demonstrating that informant discrepancies are more than measurement error: Toward guidelines for integrating data from multi-informant assessments of youth mental health. *Journal of Clinical Child & Adolescent Psychology*, *51*(1), 1–18. <https://doi.org/10.1080/15374416.2022.2158843>.
- De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. (2013). Principles underlying the use of multiple informants' reports. *Annual Review of Clinical Psychology*, *9*, 123–149. <https://doi.org/10.1146/annurev-clinpsy-050212-185617>.
- 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*(4), 858–900. <https://doi.org/10.1037/a0038498>.
- Evans, S. C., Cooley, J. L., Blossom, J. B., Pederson, C. A., Tampke, E. C., & Fite, P. J. (2020). Examining ODD/ADHD symptom dimensions as predictors of social, emotional, and academic trajectories in middle childhood. *Journal of Clinical Child & Adolescent Psychology*, *49*(6), 912–929. <https://doi.org/10.1080/15374416.2019.1644645>.
- Frazier, T. W., Youngstrom, E. A., Glutting, J. J., & Watkins, M. W. (2007). ADHD and achievement: Meta-analysis of the child, adolescent, and adult literatures and a concomitant study with college students. *Journal of Learning Disabilities*, *40*(1), 49–65.
- Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C., & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*, *98*(1), 29–43. <https://doi.org/10.1037/0022-0663.98.1.29>.
- Gerhardstein, R. R., Lonigan, C. J., Cukrowicz, K. C., & McGuffey, J. A. (2003). Factor structure of the Conners' teacher rating scale-short form in a low-income preschool sample. *Journal of Psychoeducational Assessment*, *21*(3), 223–243.
- Ginsburg, H., & Baroody, A. J. (2003). *Test of Early Mathematics Ability, 3rd Ed* (TEMA-3). Austin, TX: Pro-Ed.

- Gray, S. A., Dueck, K., Rogers, M., & Tannock, R. (2017). Qualitative review synthesis: The relationship between inattention and academic achievement. *Educational Research*, 59(1), 17–35. <https://doi.org/10.1080/00131881.2016.1274235>.
- Holmberg, K., & Bölte, S. (2014). Do symptoms of ADHD at ages 7 and 10 predict academic outcome at age 16 in the general population? *Journal of Attention Disorders*, 18(8), 635–645. <https://doi.org/10.1177/1087054712452136>.
- Jungersen, C. M., & Lonigan, C. J. (2020). *Agreement between parent and teacher ratings of problem behaviors: The role of children's executive function*. Florida State University, Tallahassee, FL.
- Jungersen, C. M., & Lonigan, C. J. (2021). Do parent and teacher ratings of ADHD reflect the same constructs? A measurement invariance analysis. *Journal of Psychopathology and Behavioral Assessment*, 43, 478–792.
- Justice, L. M., Mashburn, A. J., Hamre, B., & Pianta, R. C. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, 21(1), 51–68. <https://doi.org/10.1016/j.ecresq.2007.09.004>.
- Kulkarni, T., Sullivan, A. L., & Kim, J. (2020). Externalizing behavior problems and low academic achievement: Does a causal relation exist? *Educational Psychology Review*, 33(3), 915–936. <https://doi.org/10.1007/s10648-020-09582-6>.
- LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods*, 17(4), 433–451.
- Lane, B. R., Paynter, J., & Sharman, R. (2013). Parent and teacher ratings of adaptive and challenging behaviours in young children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 7(10), 1196–1203. <https://doi.org/10.1016/j.rasd.2013.07.011>.
- Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2002). *Preschool Comprehensive Test of Phonological and print Processing*. Authors.
- Lonigan, C. J., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2007). *Test of Preschool early literacy*. ProEd.
- Martel, M. M., Eng, A. G., Bansal, P. S., Smith, T. E., Elkins, A. R., & Goh, P. K. (2021). Multiple informant average integration of ADHD symptom ratings predictive of concurrent and longitudinal impairment. *Psychological Assessment*, 33(5), 443–451. <https://doi.org/10.1037/pas0000994>.
- Massetti, G. M., Lahey, B. B., Pelham, W. E., Loney, J., Ehrhardt, A., Lee, S. S., & Kipp, H. (2008). Academic achievement over 8 years among children who met modified criteria for attention-deficit/hyperactivity disorder at 4–6 years of age. *Journal of Abnormal Child Psychology*, 36(3), 399–410. <https://doi.org/10.1007/s10802-007-9186-4>.
- Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradović, J., Riley, J. R., & Tellegen, A. (2005). Developmental cascades: Linking academic achievement and externalizing and internalizing symptoms over 20 years. *Developmental Psychology*, 41(5), 733–746. <https://doi.org/10.1037/0012-1649.41.5.733>.
- Narad, M. E., Garner, A. A., Peugh, J. L., Tamm, L., Antonini, T. N., Kingery, K. M., Simon, J. O., & Epstein, J. N. (2015). Parent-teacher agreement on ADHD symptoms across development. *Psychological Assessment*, 27(1), 239–248. <https://doi.org/10.1037/a0037864>.
- Nelson, J. M., & Harwood, H. R. (2011). A meta-analysis of parent and teacher reports of depression among students with learning disabilities: Evidence for the importance of multi-informant assessment. *Psychology in the Schools*, 48(4), 371–384. <https://doi.org/10.1002/pits.20560>.
- Ogg, J., Volpe, R., & Rogers, M. (2016). Understanding the relationship between inattention and early literacy trajectories in kindergarten. *School Psychology Quarterly*, 31(4), 565–582. <https://doi.org/10.1037/spq0000130>.
- Owens, J. S., Storer, J., Holdaway, A. S., Serrano, V. J., Watabe, Y., Himawan, L. K., Krelko, R. E., Vause, K. J., Herrera-Girio, E., & Andrews, N. (2015). Screening for social, emotional, and behavioral problems at kindergarten entry: Utility and incremental validity of parent report. *School Psychology Review*, 44(1), 21–40.
- Owens, J. S., Holdaway, A. S., Serrano, V. J., Himawan, L. K., Watabe, Y., Storer, J., Grant, M., & Andrews, N. (2016). Screening for social, emotional, and behavioral problems at kindergarten entry: The utility of two teacher rating scales. *School Mental Health*, 8, 319–331. <https://doi.org/10.1007/s12310-016-9176-1>.
- Purpura, D. J., & Lonigan, C. J. (2009). Conners' teacher rating scale for preschool children: A revised, brief, age-specific measure. *Journal of Clinical Child and Adolescent Psychology*, 38(2), 263–272. <https://doi.org/10.1080/15374410802698446>.
- Romero-López, M., Quesada-Conde, A. B., Bernardo, G., & Justicia-Arráez, A. (2017). The relationship between executive functions and externalizing behavior problems in early childhood education. *Procedia-Social and Behavioral Sciences*, 237, 778–783. <https://doi.org/10.1016/j.sbspro.2017.02.121>.
- Starkey, P., Klein, A., & Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, 19(1), 99–120. <https://doi.org/10.1016/j.ecresq.2004.01.002>.
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245–251.
- 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(3), 314–320. <https://doi.org/10.1007/s10862-013-9348-3>.
- Swanson, H. L. (2011). Working memory, attention, and mathematical problem solving: A longitudinal study of elementary school children. *Journal of Educational Psychology*, 103(4), 821–837. <https://doi.org/10.1037/a0025114>.
- Tarver, J., Daley, D., & Sayal, K. (2014). Attention-deficit hyperactivity disorder (ADHD): An updated review of the essential facts. *Child: Care Health and Development*, 40(6), 762–774. <https://doi.org/10.1111/cch.12139>.
- Wang, L. J., Chen, C. K., & Huang, Y. S. (2015). Neurocognitive performance and behavioral symptoms in patients with Attention-Deficit/Hyperactivity disorder during twenty-four months of treatment with methylphenidate. *Journal of Child and Adolescent Psychopharmacology*, 25(3), 1557–8992. <https://doi.org/10.1089/cap.2014.0015>.
- Yang, Y., Shields, G. S., Zhang, Y., Wu, H., Chen, H., & Romer, A. L. (2022). Child executive function and future externalizing problems: A meta-analysis of prospective longitudinal studies. *Clinical Psychology Review*, 97. <https://doi.org/10.1016/j.cpr.2022.102194>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.