California State University, Monterey Bay

Digital Commons @ CSUMB

Education and Leadership Faculty Publications and Presentations

Education and Leadership

10-2022

Academic Achievement and Relations to Externalizing Behavior: Much Ado About Nothing?

Tara Kulkarni

Amanda L. Sullivan

Follow this and additional works at: https://digitalcommons.csumb.edu/educ_fac

This Article is brought to you for free and open access by the Education and Leadership at Digital Commons @ CSUMB. It has been accepted for inclusion in Education and Leadership Faculty Publications and Presentations by an authorized administrator of Digital Commons @ CSUMB. For more information, please contact digitalcommons@csumb.edu.

Journal of School Psychology 94 (2022) 1–14

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Academic achievement and relations to externalizing behavior: Much ado about nothing?

Tara Kulkarni [a,*,1], Amanda L. Sullivan [b]

[a] *California State University, Monterey Bay, USA*
[b] *University of Minnesota, USA*

ARTICLE INFO

Editor: Craig A. Albers
Action Editor: Nicholas Benson

Keywords:
Externalizing behavior
Academic achievement
Child development
Racial disparities

ABSTRACT

Some of the worst long-term outcomes of children are associated with the presence of externalizing behavior and low academic achievement. However, the nature of the causal and predictive relationship between these two domains remains contested due to inconsistent findings in previous literature. Leveraging a nationally representative sample ($N = 7330$) from the Early Childhood Longitudinal Study (ECLS)–Kindergarten Cohort of 2011, we used latent class growth analysis and machine learning cross validation techniques to analyze the association of early math and reading achievement with the development of externalizing behavior trajectories in elementary school. Several theoretically and empirically important covariates were utilized to develop a profile of learners in each trajectory. Results indicated stable teacher ratings of behavior across kindergarten to fifth grade and three primary trajectories, consisting of (1) higher persistent, (2) low persistent, and (3) no problem behavior. Importantly, teacher rated early inattention and approaches to learning behaviors, rather than direct standardized measures of academic achievement, were the strongest malleable predictors to trajectory membership. Student demographics, including being a boy and identifying as Black, contributed to these students being almost twice as likely to belong to the higher problem behavior trajectory as compared to girls and White peers. Educational implications for intervention, as well as the influence of implicit bias and structural racism in the role of teacher ratings, are discussed.

The notion that PreK–12th grade students' academic and externalizing behavior outcomes are linked is so ingrained in the minds of educators (Algozzine et al., 2011) that not only is it the basis of several interventions (e.g., Gettinger et al., 2021), but it also informs various public policies (e.g., Every Student Succeeds Act, 2015). Consequently, several states and numerous school systems have adopted small group interventions (Jones et al., 2011), universal evidence-based social-emotional learning programs (Collaborative for Academic, Social, and Emotional Learning (CASEL), 2013), and school-wide positive behavioral supports (SWPBIS; Horner & Sugai, 2015) to improve students' behavioral and academic outcomes. Although evidence on the causal, and even predictive, associations between externalizing behavior and academic achievement is mixed at best (Algozzine et al., 2011), there is an added common assumption that the association is the same within the entire student population (Kulkarni et al., 2020). This assumption could unintentionally mask differential group mechanisms (Kazdin, 2007).

* Corresponding author at: 100 Campus Center, Building 3, Sea Side, CA 93955, USA.
E-mail address: tkulkarni@csumb.edu (T. Kulkarni).

[1] Author Note: Department of Education and Leadership, California State University, Monterey Bay; Amanda L. Sullivan, Department of Educational Psychology, University of Minnesota.

https://doi.org/10.1016/j.jsp.2022.07.004
Received 11 August 2021; Received in revised form 23 April 2022; Accepted 22 July 2022
Available online 11 August 2022
0022-4405/© 2022 The Authors. Published by Elsevier Ltd on behalf of Society for the Study of School Psychology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Given that current evidence identifies at least three different group trajectories of externalizing behavior (e.g., Olson et al., 2017), consideration of how associations differ across trajectories of externalizing behavior could support a more accurate representation of how early academic achievement is associated with the development of externalizing behavior, as well as inform subsequent practice, programming, and policy. Therefore, the primary purpose of this study was to apply person-centered analyses to (a) identify if membership to trajectories of externalizing behavior may be differentially associated with early math and reading achievement, and (b) describe child, family, and school characteristics, if any, that are related to membership to those trajectories.

1. Summary of literature

Children with externalizing behavior problems are characterized primarily by aggressive, disruptive, and hyperactive behavior (Achenbach & Rescorla, 2001). Those identified to have externalizing behavior problems in kindergarten have some of the poorest educational outcomes, including lower rates of academic engagement, poor student-teacher relationships (Sabol & Pianta, 2012), and higher rates of school dropout (Piquero et al., 2012). Externalizing problem behavior was thought to remain largely stable through elementary school (e.g., Campbell et al., 2006); however, recent research has identified multiple developmental trajectories in which children's initial levels of problem behavior may either worsen, improve, or even decrease to non-clinical levels over the course of elementary school (e.g., Olson et al., 2017).

The seminal theoretical literature on the development of trajectories of externalizing problem behavior was based on Moffitt's (1993) review of the development of antisocial behavior in boys from early and middle childhood to adulthood. Although antisocial behavior is not perfectly aligned with what is currently conceptualized as externalizing behavior (Achenbach & Rescorla, 2001), Moffit's developmental theory was the first that identified multiple trajectories of problem behavior. Moffit conceptualized a 3-class taxonomy consisting of (a) life course persistent problem behavior, (b) problem behavior that begins in adolescence, and (c) low to no antisocial behavior. Subsequent research has been supportive of differential trajectories of externalizing behavior. It consistently points to the existence of three to six distinct trajectories, depending on the sample and predictors used. For example, Kjellstrand et al. (2018) used a sample of children with incarcerated parents and found four distinct trajectories, whereas Figge et al. (2018) used a sample from the Longitudinal Studies in Child Abuse and Neglect and found five distinct trajectories. Other studies, utilizing similar samples, have found three trajectories (e.g., Yoon et al., 2017). Although the number of trajectories has varied, studies have consistently found that the majority of children show low frequencies of externalizing behavior, and a smaller percentage of children belong to a high externalizing behavior trajectory. Membership to this group is associated with several negative outcomes, including low academic achievement, poor peer relations, and longer-term mental health challenges (e.g., Reef et al., 2011; Zimmermann et al., 2013).

The development of externalizing behavior difficulties can be conceptualized through an ecological systems' lens in which the behavior results from a complex interaction of individual, family, and broader systemic factors (Bronfenbrenner, 1994; Li et al., 2017). Most developmental literature in this domain has focused on child and family characteristics (e.g., coercive parenting styles), adverse childhood experiences, and maternal stress as antecedents to at-risk behavior (Girard et al., 2019). Fewer studies have examined broader systems (e.g., schools, neighborhoods); however, lower feelings of school safety and exposure to neighborhood violence are two factors that have been identified as predictors of higher externalizing behavior difficulties (Figge et al., 2018). More broadly, scholars have pointed to the challenges of differentiating intraindividual difference from racialized and gendered school contexts that can nurture externalizing presentations, or at least adults' perceptions thereof (e.g., Sullivan, 2017). Yet these considerations notwithstanding, academic achievement and its relation to the development of externalizing behavior has also been studied extensively, although consensus on how they are related has remained elusive (Kulkarni et al., 2020).

The mechanisms underpinning causal and predictive relations between early academic failure and development of behavior problems have been theorized in several ways. Early developmental theories stated that low self-esteem from academic failure could result in children communicating their frustration in the form of externalizing behavior (e.g., Dishion et al., 1991). Similarly, behavioral frameworks suggest that work-refusal to avoid the challenging task, in the form of externalizing behavior, could be a path by which academic difficulties develop into problem behavior (McIntosh et al., 2006). Bidirectional or reciprocal effects have also been hypothesized; for instance, children entering school with disruptive problems or academic difficulties may face peer and teacher rejection or disciplinary consequences, and the resultant feelings of failure, low engagement, and loss of instructional time work in parallel to worsen both academic and behavioral outcomes (Dishion et al., 1991). Lastly, some theorize that common antecedents, like language difficulties, result in the observed comorbidity (Trzesniewski et al., 2006).

There is a robust literature that highlights the comorbidity between higher externalizing behavior problems and lower academic achievement (Algozzine et al., 2011). Additionally, intervention studies integrating both academic and behavioral components have been successful at improving academic achievement as well as behavioral outcomes (e.g., Gettinger et al., 2021). Still, the empirical evidence backing predictive and causal theories is mixed. Hinshaw (1992) conducted the first review that examined the reciprocal causal effects achievement and externalizing behavior may have on each other. Hinshaw (1992) reported that, overall, although some studies reported small predictive associations between externalizing behavior and academic achievement in both directions, evidence pointed to antecedent variables, for instance, inattention, that were likely contributing to any observed covariance. Additionally, methodological limitations prevented any conclusions about causality. More recently, Kulkarni et al. (2020) conducted a systematic review on the same topic and came to similar conclusions. Most of the studies reviewed found no meaningful effects between the two domains (e.g., Duncan et al., 2007). However, some researchers reported small to moderate negative effects in the direction of early behavioral difficulties and lower academic achievement in middle and high school (e.g., van Lier et al., 2012), whereas a smaller literature base proposed that significant negative associations exist in the opposite direction (i.e., early academic difficulties are associated with later externalizing behavior; for example, see Burt & Roisman, 2010). Importantly, findings from the review highlighted the role statistical design and covariates played in contradictory results (Kulkarni et al., 2020), hampering the ability to draw conclusions that could inform practice.

Non-significant findings may be due to theoretical antecedent variables being associated with both achievement and externalizing behavior, thus providing evidence to support an antecedent model (Hinshaw, 1992). The most common confound has been inattention, which has been reported to account for comorbidity and even predictive associations between achievement and behavior (Metcalfe, Harvey, & Laws, 2013; Rennie et al., 2014). Additionally, language impairment recently has been spotlighted as a possible antecedent to both academic difficulties (Peng, Namkung, Barnes, & Sun, 2016) and behavioral problems (Chow et al., 2018). Differential conclusions regarding the predictive and causal relations between achievement and externalizing behavior have also theorized to be, in part, due to heterogeneous study samples, measurement error in assessments used, and inadequate statistical controls. For example, socioeconomic status, gender, and race may have a moderating effect on the association between the two domains (e.g., Kremer et al., 2016; Lam, 2014), pointing to the importance of context. Additionally, past scholarship has largely relied on variable centered approaches that assume the relationship is homogenous within the population, rather than the differential development of achievement and behavior within the individual (Kulkarni et al., 2020). Taken together, this body of work raises questions regarding the common presumption in policy and programming that performance in these two domains is causally related, which may, in turn, help explain mixed and null effects of various schoolwide social-emotional and behavioral support frameworks on students' academic performance (e.g., Corcoran et al., 2018; Noltemeyer et al., 2019).

Given this complexity and challenges in building an evidence base for causality with psychological constructs (Schneider et al., 2007) like achievement and behavior, it is not surprising that conclusive evidence remains elusive. However, validating causal associations is integral to scientific fields interested in improving outcomes and evaluating the effectiveness of interventions (Glass et al., 2013). This is because designing interventions that manipulate variables that are just correlates and have not been established to have temporal precedence or direct predictive associations have little applied use (Kraemer et al., 2001). Specifically, despite a lack of consensus on the predictive association between achievement and externalizing behavior (Hinshaw, 1992; Kulkarni et al., 2020), correlations between these domains have been assumed to have a causal link not only by practitioners but research bodies as well (e.g., CASEL, 2021). Given the negative outcomes associated both with externalizing behavior problems and with low achievement, a valid representation of the relation becomes critical to informing prevention efforts (McIntosh et al., 2006). We attempted to fill some of these gaps by using a person-centered analysis, which is useful in identifying developmental patterns, and group differences therein, among individuals (Laursen & Hoff, 2006). Application of the latter, with incorporation of appropriate controls and a nationally representative dataset, allows for addressing some of the key limitations of the prior work in this literature.

2. Person- versus variable-centered approaches

In most classic developmental literature, it has been the variable or construct that has been of primary interest, and how this variable operated in relation to others across the population of interest (e.g., Masten et al., 2005). The underlying assumption of these variable-centered approaches is that a homogeneity exists in the population in terms of how a variable (e.g., externalizing behavior) might behave across a population (Bates, 2000). Thus, research questions would focus on how variables were related to each other and how they developed over time using methods like multiple regression or structural equation modeling. Conversely, person-centered analyses make no assumptions about homogeneity and allow researchers to model unobserved heterogeneity in the growth of a specific variable across individuals (Laursen & Hoff, 2006). This approach is especially suited to examining psychological constructs for which the assumption that a single or average trajectory is representative of an entire population can mask heterogenous development of a characteristic across time (Suzuki et al., 2021). Specifically, person-centered approaches "describe[e] differences among individuals in how variables are related to each other" (Laursen & Hoff, 2006, p. 379) and thus are more suited to model individual development, as these approaches allow for the identification of different trajectories of development and of a profile of individuals who are most likely to belong to a category based on several characteristics. Latent class growth analysis (LCGA) and growth mixture modeling (GMM) are examples of person-centered approaches that have been used extensively to model externalizing behavior (Nagin et al., 2018).

These methods can help researchers identify whether academic achievement and externalizing behavior share different relationships based on groups within a population. For example, we could examine specifically whether improvement in academic achievement can support improvement in behavior for those with the highest frequencies of externalizing behavior problems and whether this relationship is different for those with lower levels of externalizing behavior. This is critical given that mixed results in previous scholarship may mask legitimate relationships due to the assumption of homogeneity. Additionally, person centered analyses allow researchers to adopt a multi-dimensional representation of individuals, acknowledging intersectional identities, as well as capturing nuance in model outcomes (Suzuki et al., 2021). However, applications of person-centered approaches in the literature primarily have been focused on the association of family characteristics with the development of externalizing behavior (e.g., see Olson et al., 2017) rather than academic achievement.

3. Research purpose

This study was motivated by three research needs. First, mental health or behavior prevention and intervention requires substantial investment because of the costs of materials, personnel, time, and evaluation (Belfield et al., 2015). It is crucial for intervention developers and educational leaders to have an accurate representation of the nature of construct relations and intervention efforts and potential outcomes. In this case, whether intervening on academic achievement can indeed result in improved behavioral outcomes or vice versa. Second, identifying whether this relationship is stronger for certain subgroups of students is critical to inform prevention efforts, especially in the early elementary years, as schools could target intervention efforts on one domain to improve both academic and behavioral functioning (Masten & Cicchetti, 2010), and broader outcomes shown to be related to difficulties (van Lier et al., 2012). Finally, the history of contrary findings on this issue contributes to the replication crisis in psychology (Maxwell et al., 2015) and this study contributes to this body of literature, leveraging a nationally representative longitudinal sample, appropriate controls, and a rigorous research design. The research questions in this study included the following:

1. Are students' early math and reading achievement associated with membership to differential externalizing behavior trajectories through elementary school?
2. Which child and family characteristics are associated with membership to trajectory classes?

It was hypothesized that academic achievement would have a stronger relationship with the highest frequency trajectory of externalizing behavior. The second research question acknowledges that, like most psychological concepts, behavior and achievement have several correlates that could be related to both. We considered a range of child and family characteristics in an effort to elucidate competing theories of the relations of academic and behavioral performance (e.g., psychological antecedents, contextual factors). It was hypothesized that a student's language development (Chow et al., 2018), and inattentive behaviors (Rennie et al., 2014) would be related to both achievement and behavioral outcomes. Similarly, a child's access to resources at home and in school have been previously associated with the development of their academics and behavior (Lam, 2014). Thus, examining how early achievement is associated with the trajectories of behavior without accounting for these variables could lead to an inaccurate representation of how student behavior develops in schools (Kamata et al., 2018).

4. Method

This study utilized the Early Childhood Longitudinal Study–Kindergarten Cohort of 2011 (ECLS-K:2011) collected by the U.S. Department of Education's National Center for Education Statistics. The Early Childhood Longitudinal Study has several advantages over other nationally representative datasets. It is the most current and therefore representative of the current cohort of elementary students in the United States, and it was designed to provide multi-informant quantitative information on the progress of children's family, health, behavioral, and cognitive status through elementary school (Tourangeau et al., 2018). To allow for longitudinal analysis, assessments and instruments used through the grades were largely the same across waves of data collection. The ECLS-K:2011 used a multistage, clustered sampling design to allow precise estimates for the population within the data. From the original 90 primary sampling units (PSUs), 30 were selected for data collection starting in the fall of kindergarten. The final sample of students (n = 18,170) was followed up to fifth grade unless the child died or left the United States.

For the present study, we used 2010–2011 child level data collected in the spring of kindergarten, first, second, third, fourth, and fifth grades. The ECLS authors provided several weights and their corresponding PSU and strata for researchers to use in order for the data to remain nationally representative. The weights accounted for attrition, missingness, and the clustered nature of the data, and researchers selected a weight in order to retain the maximum number of observations (Tourangeau et al., 2018). The specific weight used was W9C19P_2T290 and it resulted in an unweighted analytic sample of 7330 students (all N's rounded to 10 as per NCES requirements). Bar plots, correlation matrices, and Mplus code are available in the Supplementary Materials.

4.1. Measures

We used teacher ratings to construct behavioral measures because the behaviors of interest were in the school setting. The externalizing behavior subscale ($\alpha = 0.87$–0.89), which was a modified version of the Social Skills Rating System (SSRS; Gresham & Elliott, 1990) designed for the ECLS, was administered in the spring of kindergarten, first, second, third, fourth, and fifth grades and was used as the longitudinal outcome in this study. The SSRS has a history of strong construct validity across gender, ethnicities, and international populations (Barger et al., 2022; Matson & Wilkins, 2009). Additionally, the SSRS has shown moderate to strong concurrent validity with the Behavior Assessment Scale for Children (Lane et al., 2019; Reynolds & Kamphaus, 2015), which is the most used assessment of behavior in school psychology (Benson et al., 2019). It is important to note that the SSRS has been validated for use as a screener and not as a diagnostic measure (Sullivan et al., 2021), thus trajectories are not diagnostic in nature, but rather a representation of students who teachers consider to be at risk. Six items were included in the teacher questionnaire. Items in the SSRS ask teachers if the "child disobeys rules", "fights with others", "has temper tantrums", "is aggressive towards people or objects", "does things to make others scared", and "talks at times when one is not supposed to" (Gresham & Elliott, 1990). The variable presented in the dataset is the mean of the summed of scores on this scale (1 = *Never* to 4 = *very often*). If a teacher marked the option *no opportunity to observe*, this was counted as missing data. Kindergarten and first grade spring subscales consisted of 5 items, whereas second, third, fourth, and fifth grade subscales consisted of 6 items.

4.1.1. Academic achievement

Math and reading theta scores (range = −4, 4) for each child were calculated based on direct assessments given to children in each wave of data collection. Reading and math achievement in fall kindergarten ($\alpha = 0.95$ and 0.94, respectively) were the primary predictors in this analysis. Items on both scales were based on the 2009 reading and math frameworks for the National Assessment of

Education Progress (Tourangeau et al., 2018).

4.1.2. Behavioral covariates

As covariates, we used teacher reported measures of approaches to learning and inattention in spring of kindergarten given previous research suggesting that both could act as confounds of the relationship between externalizing behavior and achievement (Gray et al., 2017). The Approaches to Learning measure ($\alpha = 0.91$) was a 7-item questionnaire developed for the ECLS based on the dimensions of school readiness described by the National Education Goals Panel framework (NEGP; Kagan et al., 1995) to encapsulate classroom-based behaviors considered conducive to learning in a school environment. Items include "keeps belongings organized", "shows eagerness to learn new things", "works independently", "easily adapts to changes in routine", "persists in completing tasks", "pays attention well", and "follows classroom rules". The Inattention measure ($\alpha = 0.87$) was a 12-item subscale of the short form Children's Behavior Questionnaire (CBQ; Putnam & Rothbart, 2006) administered in the spring of kindergarten. The CBQ asks teachers to score how often children exhibited behavior on a 7-point scale (1 = *extremely untrue* to 7 = *extremely true*). Altough example items of the CBQ items are not presented due to copyright agreements, items were designed to measure behaviors that demonstrate a child's ability to focus their attention on tasks relevant to the classroom environment (Putnam & Rothbart, 2006).

4.1.3. Working memory and language covariates

The analysis included as a covariate a direct assessment of working memory where children were asked to complete a backward-digits task that required the child to orally repeat numbers in the reverse order that was presented to them. Additionally, for early language skills, the ECLS included scores from two tasks from the Preschool Language Scale (Duncan & DeAvila, 1998). The "Simon Says" task required students to follow simple spoken instructions in English and the "Art Show" task was a picture vocabulary assessment (Tourangeau et al., 2018). The combined score of the language screener in the spring of kindergarten was used in the analysis ($\alpha = 0.89$). These two covariates were measured in the fall and thus were considered antecedent variables in the analysis.

4.1.4. Sociodemographic covariates

The analysis included a child's sex, race, family socioeconomic status, parent's educational attainment, and percentage of non-white students in school to account for associations between sociodemographic indicators, including school segregation (Palardy et al., 2015), on externalizing problem behavior and academic achievement. When including race as a variable in the analysis, it is important to acknowledge that it is not measuring biological differences between groups, but rather differences in how children are treated due to their race within school systems, including by teachers (Gillborn et al., 2018). This is especially relevant in this study as ratings of behavior were completed by teachers.

4.2. Data analysis

The ECLS developers provided an average of the externalizing behavior subscale in the dataset based on the summed frequency of behaviors by number of items, resulting in a variable that mimicked a continuous variable. However, treating it as continuous and normal could result in inaccurate standard errors (Shiyko et al., 2012). Thus, to account for the original frequency characteristic of the measure and the non-normal positive skew, the mean was multiplied by the number of items on the subscale. Although the kindergarten and first grade subscales had one less item (i.e., 5 items instead of 6 items), all means were multiplied by six to maintain comparisons with other rounds of data collection. Thus, the total sum frequency of behaviors for each student was modeled (Collins et al., 2017) instead of using their mean score on the subscale. Additionally, the ECLS authors coded the zero-frequency option in the externalizing subscale (*never shows behavior*) as '1', rather than zero. To adjust for this, each student's score was subtracted by six. As a result, children who did not show any externalizing behavior on any item on the scale had a score of zero, rather than six, allowing clearer interpretation without disturbing the properties of the original distribution. Given our large sample size, multicollinearity was not a concern and the correlation matrix confirmed that all inter-variable correlations were below 0.6.

By using the weight to select cases to include in the analysis, it was found that less than 5% of the data were missing across every time point and variable except for birth weight and IEP status, for which approximately 8% and 10% of data were missing, respectively. However, given that ECLS consists of several related variables, including parent and teacher reports of inattention, internalizing, approaches to learning, gender, sex, socioeconomic status, and birthweight, this dataset qualified for the MAR assumption that allows for Full Information Maximum Likelihood Estimation, which is a robust method to estimate missing data even with non-normal data (Larsen, 2011).

4.2.1. Growth curve modeling and mixture modeling

In cases where individuals are theorized not to come from the same population, mixture modeling provides researchers a way to model unobserved heterogeneity. The growth mixture model (GMM) traditionally allows within class variation in intercept and slope (i.e., assumes that the mean trajectory of a class is made up of multiple trajectory subpopulations; Nagin et al., 2018), whereas in the LCGA the growth curve parameters are fixed and not random (Reinecke & Seddig, 2011). In the present study, first an LCGA and then a GMM were run as per best practice recommendations (Reinecke & Seddig, 2011). However, only the former method was finally used due to non-convergence of the GMM models. This was hypothesized to be because of the complexity of GMM that often yields unreliable solutions due to non-convergence of multiple local maxima (e.g., Olson et al., 2017). The code and output for both GMM and LCGA solutions can be found in the Supplementary Materials. Finally, to account for the large number of zeros when modeling behavior, we ran an unconditional base model with a negative binomial model (NB) and a Zero Inflated Poisson (ZIP) to test which model returned the best fit statistics (Allison, 2011). We used the model with best fit to specify the distribution to run the latent class growth analysis on.

Model Building. In this analysis, we used the 1-step or direct method. We simultaneously added time invariant covariates in the longitudinal model to estimate class trajectories, eliminating (a) the assumption that class membership is concrete or known and (b) the bias of classification error that is common in the traditional 3-step method (Lanza & Cooper, 2016). Additionally, this method accounts for the common finding that covariates alter class structure and does not present artificial class trajectories that assume a person's membership to a trajectory cannot be influenced by variables theoretically related to the primary outcome (Kamata et al., 2018).

Cross Validation. Machine learning researchers suggest the use of cross validation to estimate the performance of the method used to build the final model, as well as to improve generalizability of the chosen model to new datasets (Cawley & Talbot, 2010). We used a cross validation method known as holdout, which is preferred when using larger datasets (Gao et al., 2018). In this method, the sample is split randomly in two halves; one dataset is called the *training* set and the other the *testing* set (Cawley & Talbot, 2010). First, all model building was conducted on the training set and the optimal number of classes were obtained. Then, this model was run on the testing set to see if similar fit statistics and parameter estimates were returned. No significant differences would indicate that the model generalized well and there was no overfit (Gao et al., 2018). The final analysis could then be run on the entire data set.

5. Results

We describe steps taken for model selection, examine descriptives of the final model chosen, and then conclude by examining both research questions. Descriptives of the weighted sample are presented in Table 1. Missing data percentages are presented for all covariates and the primary outcome variable at each wave. FIML, as stated earlier, was used to handle missing data and thus no cases were dropped.

5.1. Model selection

We first conducted model selection on the training dataset ($n = 3660$) as per cross validation procedures. We ran an unconditional latent curve growth model to compare the fit between a ZIP model and NB model. We used commonly accepted fit indices like the Akaike information criterion (AIC), and adjusted Bayesian information criterion (BIC) to judge fit. As smaller adjusted BIC values indicate better fit we chose the NB model (ΔBIC 536.75). Next, we chose the best fitting form of the developmental trajectory. The

Table 1
Descriptives of weighted sample.

Variable	*M*/Pct (*SD*)	Range of possible values	SE	% missing
Child Characteristics				
Male	51.43		0.007	0
White	51.71		0.023	0
Black	13.36		0.167	0
Hispanic	24.82		0.020	0
Other	10.10		0.011	0
Birth weight in pounds	6.85 (1.31)	5–8	0.021	10.89
Reading IRT K Fall	53.38 (11.31)	0–205	1.50	0.65
Math IRT K Fall	35.46 (11.44)	0–206	1.53	0.26
Numbers reversed score K Fall	98 (16.82)	45–200	3.32	0.40
Attention K	3.03 (0.88)	1–5	0.001	3.90
Approaches to Learning K	2.92 (0.67)	1–4	0.005	3.03
Internalizing symptoms K	1.50 (0.48)	1–4	0.002	1.02
Total language score K fall	18.73 (2.75)	0–20	0.11	1.02
IEP % in kindergarten spring	8.8	0–100	0.000	8.84
Percent minoritized students in school	14.64 (22.35)	0–100	6.41	0.22
Externalizing K	3.69 (3.78)	0–18	0.17	1.05
Externalizing 1st grade	4.27 (3.62)	0–18	0.16	3.69
Externalizing 2nd grade	4.27 (3.70)	0–18	0.16	0.68
Externalizing 3rd grade	4.09 (3.65)	0–18	0.16	0.60
Externalizing 4th grade	3.88 (3.57)	0–18	0.15	0.98
Externalizing 5th grade	3.82 (3.58)	0–18	0.15	1.05
Family Characteristics				
Socio-economic status	−0.09 (0.77)	−2.33-2.60	0.007	0.37
Parent Education Level	4.56 (1.82)	1–8	0.040	0.29

Note. Weighted estimates reported. Unweighted $N = 7330$. All unweighted counts rounded to 10 as per NCES regulations. Birth weight in pounds, math, and reading scores are scaled scores; the numbers reversed score is a standard score. Attention, approaches to learning, and internalizing ranges are means of rating scales where 1 = *never shows behavior*, 4/5 = *almost always shows behavior*. Externalizing range is 0 = *marked as never shows externalizing behavior across all items*, 18 = *marked as almost always shows behavior across all items*. Parent education was coded as 0 = *none*, 1 ≤ *8th grade*, 2 = *9th–12th grade*, 3 = *high school diploma*, 4 = *vocational program*, 5 = *some college*, 6 = *bachelor's degree*, 7–8 = *Master's degree or higher*. The SES variable is a continuous imputed variable with 0 representing the median average household salary in the US.

estimated and observed mean plot indicated, that on average, teacher perceived behavior was stable across time, with a slight increase in teacher-reported externalizing behavior in first and second grade that largely disappeared by fourth grade. Overall, there was little change in the mean teacher-reported externalizing behavior from kindergarten to fifth grade. We analyzed four unconditional NB latent curve growth models to account for the increase in behavior in first and second grades, including (a) quadratic, (b) piecewise, (c) linear, and (d) free loading slope factor. The free loading slope factor model accounted for non-linear trends and change in slope. It is interpreted as the change in behavior between the first and fifth time points.

The quadratic and piecewise models did not converge even after reducing the number of random start points. The free-loading factor model and the linear unconditional model performed the best and had very similar AIC and BIC values. However, although the former converged, the output warned of several parameters being fixed to avoid singularity of the matrix due to the model most likely not being identified. Thus, the linear model seemed to be the best model to represent the data. To make sure that adding covariates did not change this conclusion, we then reran the models with all the covariates and predictors (conditional model). Similar to the unconditional models, the piecewise and quadratic models did not converge, whereas the linear model converged with no warning messages. We reran the linear model with twice the number of random starting points to validate the replicated best log likelihood. This step ensures that the optimal solution was reached.

5.1.1. Selecting the appropriate number of trajectory classes

We ran models consistent with the procedures described by Jung and Wickrama (2008) for a 1-step approach. For each unconditional model, we ran the corresponding full model to make sure there were no major differences in class structure. Fit statistics for the unconditional and full models are presented in Table 2. We discarded the 4-class and 5-class solutions as they had poor entropy in the unconditional models and a zero-person class in the conditional models, both indicators of poor fit. Ideally, a lower BIC and an increased entropy should help select the best fitting model. Although the 3-class solution performed the best, the entropy was similar between the 2- and 3-class solution, indicating only slight improvement in model fit. This is supported by the non-significant LRT values. The statistical evidence, therefore, suggested that the 2- and 3-class models had similar fit statistics and fit the data equally well.

Box and Draper's (1987) observation that "all models are wrong, but some are useful" (p. 424) was relevant to decision making when choosing between these two models. A consistent finding is that the more commonly identified taxonomy is of higher persistent, low/moderate, and no problem groups across developmental periods (Van Dulmen et al., 2009). When applying the above findings to the current analysis, the 2-class model was unable to identify the higher persistent behavior group and split the sample almost in half, whereas the 3-class solution identified a higher group (27%), a no problem group (29%), and a majority low persistent group (43%), with the latter two groups representing typically functioning children. Thus, the ability to identify a higher problem group and alignment with empirical data, as well as the purpose of the study, led to the 3-class solution being chosen (Fig. 1).

Once the 3-class model had been chosen, this was fit to the testing dataset. There were no significant changes in fit statistics or class separation between the training and testing datasets, indicating that the model building process was accurate and supported generalization of the model. Thus, for the purpose of the present study, the 3-class, negative binomial model was chosen to address the research questions. Subsequently, all analyses were conducted with the complete sample to take advantage of all information in the dataset. To support descriptive analysis, multiple ANOVAS were also conducted to identify significant mean differences in predictor variables between the three trajectory classes.

5.2. Research Question 1

Means and standard deviations for each trajectory class in the 3-class model are presented in Table 3. The first research question examined whether early math and reading achievement predicted a student's externalizing behavior trajectory class membership to the higher persistent and no-problem groups, as compared to the low-persistent group. Results from latent class growth analysis (Table 4) showed that after accounting for various socioeconomic characteristics, teacher-reported approaches to learning, and teacher-reported inattention, a student's reading achievement had no association with trajectory type (i.e., a student's reading achievement was not associated with their classification to the HP; OR = 1.01, 95% CI [1.00, 1.02], $p = .273$) as compared to the low persistent (LP) group. Although the odds ratio p-value indicated that a student with lower reading scores was more likely to belong to the LP group rather than the NP group, the odds ratio itself indicated that the size of this effect was very small (OR = 0.98, 95% CI

Table 2
Fit statistics for 1-, 2-, and 3-class models.

	-Ll	Adj.BIC	%ΔBIC	LRT	Entropy
Conditional Models					
1 class	290,997.49	583,225.073		NA	NA
2 class	293,703.71	588,540.256	0.90	$p = .74$	0.85
3 class	291,555.36	584,357.993	−0.72	$p = .77$	0.82
Unconditional Models					
1 class	98,851.02	197,764.967		NA	NA
2 class	101,008.23	202,079.398	2.12	$p = .74$	0.84
3 class	99,160.14	198,400.369	−1.85	$p = .77$	0.81

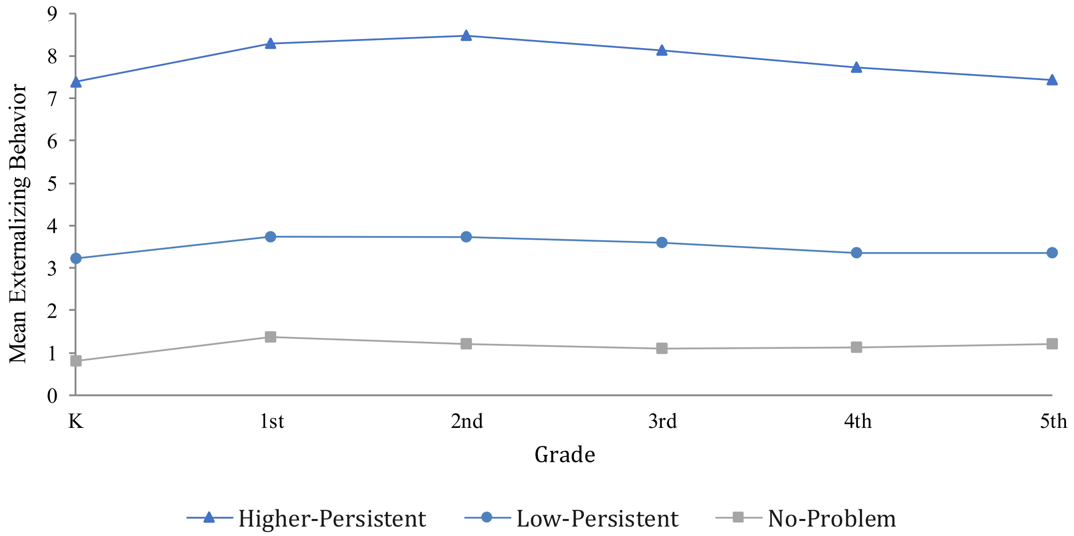

Fig. 1. *Mean* teacher ratings of externalizing behavior kindergarten through fifth grade for the 3-class *solution.*

Table 3
Mean Parameter Estimates and Classification Probabilities.

	Unweighted *N*	Intercept (SE)	Slope (SE)	CP Class 1 (Higher persistent)	CP Class 2 (No problem)	CP Class 3 (Low persistent)
Higher-persistent trajectory (27%)	2000	2.11** (0.02)	−0.02** (0.004)	0.93	0.00	0.07
Low-persistent stable trajectory (43.3%)	2160	1.29** (0.03)	−0.01* (0.006)	0.04	0.05	0.91
No-problem stable trajectory (29.5%)	3170	0.11* (0.05)	0.02 (0.01)	0.00	0.93	0.07

Note. Unweighted *N* rounded to 10 as per NCES requirements. *$p < .05$, **$p < .001$. SE = standard error.

Table 4
LCGA Regression Odds Ratio Results.

	Higher persistent stable trajectory class (HP)		No-problem stable trajectory class (NP)	
Child Characteristics	OR	95% CI	OR	95% CI
Male	2.40**	2.02, 2.86	0.48**	0.42, 0.55
Black	2.07**	1.57, 2.72	0.41**	0.28, 0.60
Hispanic	0.78*	0.65, 0.94	1.03	0.83, 1.28
Other	0.70*	0.52, 0.92	1.11	0.87, 1.41
Birth weight in pounds	1	0.95, 1.05	0.99	0.92, 0.99
Reading IRT K Fall	1.01	1.00, 1.02	0.98*	0.98, 0.99
Math IRT K Fall	1.02*	1.01, 1.03	1.01	1.00, 1.02
Numbers reversed score K Fall	1.01	1.00, 1.01	1	1.00, 1.01
Attention K spring	0.41**	0.36, 0.48	2.65**	2.31, 2.65
Approaches to Learning K spring	0.60**	0.51, 0.71	1.75**	1.50, 2.04
Internalizing symptoms K spring	1.19	0.97, 1.46	1.54**	1.31, 1.81
Total Language Score	1.03	1.00, 1.04	0.94**	0.91, 0.96
IEP in Kindergarten spring	0.9	0.73, 1.11	0.84	0.63, 1.13
Percent minoritized students in school	1	1.00, 1.01	1	1.00, 1.01
Family Characteristics				
Socio-economic status	0.60**	0.51, 0.70	1.32*	1.13, 1.54
Parent Education Level	1.06	1.00, 1.06	0.96	0.90, 1.02

Note: *$p < .05$, **$p < .001$. Reference groups are lower persistent class for no-problem and higher persistent classes, White for Black, Hispanic and Other, and female for male.

[0.98, 0.99], $p = .001$) and that students' reading achievement made them almost equally likely to belong to either group. Similarly, although significant, students' math achievement was weakly associated with classification to the HP (OR = 1.02, 95% CI [1.01, 1.03], $p = .002$) or NP group (OR = 1.01, 95% CI [1.00, 1.02], $p = .138$) as compared to the LP group. Nevertheless, class membership was significantly related to several other variables included in the model.

5.3. Research Question 2

The second research question examined the child and sociodemographic characteristics associated with externalizing behavior trajectory memberships. Boys were almost two and a half times as likely (OR = 2.42, 95% CI [2.02, 2.86], $p < .001$) than girls to belong to the higher persistent trajectory group, and almost half as likely (OR = 0.48, 95% CI [0.42, 0.55], $p < .001$) than girls to belong to the NP group as compared to the low persistent group, after controlling for covariates. As compared to White students, Black students were twice as likely (OR = 2.07, 95% CI [1.57, 2.72], $p = .002$) to belong to the HP group as compared to the LP group, and around half as likely as White students (OR = 0.41, 95% CI [0.28, 0.60], $p < .001$) to belong to the NP group as compared to the low persistent group after controlling for covariates. Hispanic students were less likely (OR = 0.78, 95% CI [0.65, 0.94], $p = .013$) than White students to belong to the HP group than the LP group and were equally likely (OR = 1.03, 95% CI [0.83, 1.28], $p = .828$) to belong to the NP group as compared to the LP group.

Socioeconomic status was also related to trajectory membership. As latent class growth analysis is essentially a multinomial regression and socioeconomic status was treated as continuous, for every one-unit increase in the socioeconomic scale the increase in odds follows a log function. Overall, students who belonged to families who were of higher socioeconomic status were less likely to be categorized in the HP group (OR = 0.60, 95% CI [0.51, 0.69], $p < .001$) and more likely to belong to the NP group than the LP group (OR = 1.32, 95% CI [1.13, 1.54], $p = .009$). Neither birthweight nor parent's education level were significantly related to class membership after controlling for covariates. Similarly, belonging to a school with a higher proportion of Black students was not associated with class membership (see Table 4). This result was replicated in the between-class descriptives (see Table 5), where students in each class belonged to schools that had similar mean percentages of minoritized students.

5.3.1. Child characteristics

The strongest malleable predictors to class membership were teacher-reported attention and approaches to learning scores in kindergarten. A student with higher ratings of teacher-reported attention had almost half the odds of belonging to the higher persistent group as compared to the low persistent group (OR = 0.41, 95% CI [0.36, 0.48]). Findings were similar for the approaches to learning

Table 5
Between Class Mean Comparison.

	Higher Persistent Group (*SD*)	No Problem Group (*SD*)	Low Persistent group (*SD*)	*F* or Chi square
Child Characteristics				
Percent boys	74.01	30.52	51.38	351.23
Percent White	47.87	55.55	51.52	6.90
Percent Black	23.14	5.38	12.63	79.46
Percent Hispanic	21.44	27.33	25.24	5.58
Percent other- race	7.52	11.75	10.61	6.34
Birth weight in pounds	6.82[a] (1.37)	6.86[a] (1.30)	6.85[a] (1.30)	1.78
Reading IRT K Fall	53.27[a] (11.08)	56.55[b] (11.86)	54.84[c] (11.07)	30.19
Math IRT K Fall	34.65[a] (10.49)	38.77[b] (11.86)	36.20[c] (11.07)	58.54
Numbers reversed K Fall	92.99[a] (16.99)	99.14[b] (16.23)	95.80[c] (16.72)	60.53
Attention K spring	2.36[a](0.79)	3.64[b](0.60)	3.05[c](0.80)	918.12
Approaches to Learning K Fall	2.61[a](0.63)	3.59[b](0.46)	3.17[c](0.60)	971.90
Internalizing K spring	1.60[a](0.54)	1.47[b](0.44)	1.46[b](0.47)	27.27
Total Language Score	18.82[a] (2.83)	18.71[a] (3.34)	18.80[a] (2.90)	1.51
Percent IEP in K spring	12.85	4.71	8.99	38.23
Externalizing K	7.39[a] (3.96)	0.81[b] (1.35)	3.23[c] (2.61)	1304.33
Externalizing 1st grade	8.30[a] (3.50)	1.37[b] (1.41)	3.74[c] (2.20)	2537.32
Externalizing 2nd grade	8.48[a] (3.48)	1.20[b] (1.29)	3.73[c] (2.24)	2252.53
Externalizing 3rd grade	8.12[a] (3.60)	1.09[b] (1.28)	3.60[c] (2.23)	2057.55
Externalizing 4th grade	7.73[a] (3.60)	1.14[b] (1.36)	3.35[c] (1.39)	2058.90
Externalizing 5th grade	7.42[a] (3.78)	1.20[b] (1.39)	3.36[c] (2.48)	2880.04
Family Characteristics				
Socio-economic status	−0.22[a] (0.72)	0.09[b] (0.82)	−.02[c] (0.77)	46.70
Parent Education Level	4.44[a] (1.68)	4.87[b] (1.94)	4.69[b] (1.87)	13.99

Note. Weighted estimates presented. All chi square and F statistics significant at $p < .01$ or $p < .001$ except for birth weight, language, and parent education level, which were not significant. Birth weight is in pounds. Math and reading scores are scaled scores. The numbers reversed score is a standard score. Attention, approaches to learning, and internalizing ranges are means of rating scales where 1 = *never shows behavior*, 4/5 = *almost always shows behavior*. Externalizing range is 0 = marked as *never shows externalizing behavior across all items*, 18 = marked as *almost always shows behavior across all items*. Parent education was coded as 0 = *none*, 1 ≤ *8th grade*, 2 = *9th–12th grade*, 3 = *high school diploma*, 4 = *vocational program*, 5 = *some college*, 6 = *bachelor's degree*, 7–8 = *master's degree or higher*. The SES variable is a continuous imputed variable with 0 representing the median average household salary in the US. Superscripts of a, b, c indicate a significant difference between means.

variable. An increase in approaches to learning teacher ratings was associated with a decreased chance of belonging to the HP group as compared to the LP group (OR = 0.60, 95% CI [0.51, 0.71], $p < .001$) and a decrease in teacher-rated approaches to learning increased the chances of a student belonging to the NP group as compared to the LP group. The working memory measure and language measure were not related to class membership after controlling for covariates. Similarly, the presence of an individualized education plan in kindergarten was also not related to class membership (Table 4). Lastly, higher ratings of teacher-reported internalizing behavior were associated with increased odds of belonging to the no problem group than the low problem group.

6. Discussion

In this study we used a person-centered latent class growth analysis to study the association of a student's early academic achievement with membership to differential trajectories of externalizing behavior in elementary school. Additionally, student and family characteristics associated with trajectory membership were examined. Overall, three class trajectories of externalizing behavior were identified, including a (a) higher persistent group (HP), (b) low persistent group (LP), and (c) no problem group (NP) based on teacher ratings. Importantly, neither a child's early math nor reading achievement was associated with trajectory membership. However, being a Black boy from a family considered to be lower socioeconomic status disproportionately increased a child's chances of belonging to the higher problems group based on teacher ratings. Additionally, children who were rated by their teacher as having higher levels of inattention and lower teacher-reported approaches to learning were all associated with classification in the HP as compared to the LP group.

6.1. Early academic achievement and externalizing behavior

Despite small significant differences for means in reading and math student achievement across groups, the latent class growth analysis finding implies that the variation in mean academic achievement at school entry observed across behavior trajectory groups was likely attenuated by the presence of other covariates. Although there are developmental (e.g., Dishion et al., 1991) and behavioral theories (McIntosh et al., 2006) that directly link low academic achievement to the development of externalizing behavior, results from this study do not support predictive relations between the two domains. This is consistent with a recent systematic review that suggested no substantial predictive relations between the two domains (Kulkarni et al., 2020); the present study generalizes these findings to different trajectories of externalizing behavior. Additionally, results lend support to Hinshaw's (1992) antecedent model theory in which the observed comorbidity between lower academic achievement and externalizing behavior is explained by theoretical antecedent variables.

Given that inattention strongly predicts academic achievement when externalizing behavior is controlled for (e.g., Gray et al., 2017) and that children in the HP group also had the lowest attention scores, it could be that inattentive symptoms, rather than level of externalizing behavior, is associated with the observed depressed academic achievement. Like inattention, low teacher ratings in approaches to learning were associated with membership to the higher persistent group as compared to the low persistent group. The measure used in this study was developed as a dimension of school readiness that encompasses behavior critical to classroom success, such as transitioning and task persistence. Possibly, students who have higher externalizing behavior in kindergarten display a combination of other problem behaviors (e.g., inattention, poor approaches to learning) that contributes to the consistent ratings of externalizing behaviors through elementary school. Finally, although recent longitudinal meta-analytic results examining the relationships between language and behavior are of small but negative associations (e.g., Chow et al., 2018), the findings from this study did not support this trend. A potential reason could be that most students performed very well on the language measure used in ECLS (Table 1), thus resulting in a lack of variation and power to predict membership to the externalizing behavior groups.

6.2. Sociodemographic characteristics

There were several student and family characteristics associated with trajectory membership. However, of most concern was that Black students' behavior ratings by teachers, especially boys, made them almost twice as likely as their White peers to belong to the HP class trajectory. This result is not insubstantial as the HP group had a mean that is at least one standard deviation above the low persistent group and almost two standard deviations above the no problem group. The perception of Black students having higher problem behaviors is multifaceted and can be viewed through an ecological lens that acknowledges multiple spheres of influence on the development and perception of behavior (Bronfenbrenner, 1994). At the student level, studies have shown that White teachers tend to rate Black students worse than other race peers when rating externalizing behavior, regardless of overall class behavioral context (e.g., Bates & Glick, 2013).

Negative racial stereotypes also make it more likely that teachers pathologize patterns of behavior in Black children as more problematic (Okonofua & Eberhardt, 2015). Furthermore, teacher ratings cannot be seen divorced from school context. Children in schools, especially Black children, are more likely to experience disproportionate levels of harsh discipline (Smolkowski et al., 2016) and higher student-teacher conflict (Zimmermann, 2018). At the school-level, due to decades of segregation and racist housing policy, Black students are more likely to attend schools with fewer resources in lower socioeconomic neighborhoods (Owens et al., 2016). These schools tend to have teachers that are poorly trained and highly stressed (Simon & Moore Johnson, 2015), resulting in conditions that exacerbate implicit bias and activation of harmful stereotypes (Langer, 2009). A large body of work suggests that educators apply harsher and more frequent penalties to Black students' behavior but that differences in behavioral outcomes (e.g., exclusionary discipline) are due to racial bias—including racially biased adult responses to normative child and adolescent behavior (Amemiya et al., 2020)—rather than differences in students' behavior (e.g., Rocque & Paternoster, 2011; Skiba, 2015). Taken together, the results of the present study should not be taken to suggest that Black boys *have* more problematic behavior but rather they are *rated* as such by teachers. The exact nature of those ratings—both in research and practice contexts—should be subject to scrutiny.

6.3. Limitations

This study has certain limitations related to study design and methodology through which results should be interpreted. Even though it leveraged a nationally representative sample, the sample size for Asian, Native American, and Pacific Islander students was not large enough to allow disaggregated analysis. Thus, results may not generalize to students who belong to these racial and ethnic categories. Second, this study only examined students in elementary school, and there is emerging evidence that early academic achievement could be related to externalizing behavior in adolescence (Weeks et al., 2016), thus future research should include middle and high school students from appropriate datasets when available. Furthermore, the latent class growth analysis in this study was conducted via M*plus* (Muthén & Muthén, 2018) and thus the analysis was constrained within its limitations. As a result, the Bootstrapped Likelihood Ratio Test (Feng & McCulloch, 1996), a recommended fit statistic, was not available for complex samples which limited the ability to use it to compare the 1-, 2-, and 3-class models. Finally, although the intention of the study was to use early childhood measures to ensure temporal precedence, results should be interpreted within the limitation of measures that are collected when children were in kindergarten.

6.4. Implications for practice

Several implications can be drawn from the results of this study. The results suggest that a student's early academic achievement does not make them more or less likely to be rated as having externalizing behaviors. Instead, what is directly related to the development of externalizing behavior are other malleable positive behaviors like attention and approaches to learning. Additionally, it could be these malleable factors are antecedent causes to both externalizing behavior and academic achievement (Hinshaw, 1992). Thus, any efforts targeting only academic achievement, whether universal or intensive, are unlikely to have an effect on decreasing or preventing the development of externalizing behavior in children. A blended model of support that targets both behavioral and academic needs of students at all levels of need is necessary to support whole-child outcomes (Horner & Sugai, 2015). In other words, the prevention of behavioral problems should depend on mental health services rather than expecting gains in one domain to carry over to the other.

Second, results support evidence that teacher-reported externalizing behavior is largely stable through elementary school. This underscores the importance of early intervention to prevent behaviors from developing or continuing. Interventions targeted at improving early school readiness skills have been shown to prevent the development of later externalizing behavior problems (Dodge et al., 2017), and recent studies have shown that interventions focusing on developing self-regulation skills, which are closely related to both attentional control and approaches to learning, could result in greater reduction of problem behavior (Duncan et al., 2018a, 2018b).

Lastly, the stability of teacher-reported behavior of students could also have implications for school systems that serve students of color, and especially Black students. Results from this study imply that negative perceptions of Black students' behavior start early and remain largely stable, supporting other findings that have found that biased observations of behavior can begin as early as preschool (e.g., Gilliam et al., 2016). Furthermore, implicit bias training may not be enough to change behavior. Rather, researchers suggest implementing cultural competency training for teachers to provide context to students' behaviors and prevent misinterpretation (Blake et al., 2016). Finally, when addressing disproportionate representation in any domain, it is important that educators do not forget that true prevention cannot focus only on the individual. Rather, to prevent inequitable outcomes that are a result of historical and structural racism, educators could instead scrutinize and dismantle the institutions that create them (Artiles et al., 2010).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsp.2022.07.004.

References

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

Algozzine, B., Wang, C., & Violette, A. S. (2011). Reexamining the relationship between academic achievement and social behavior. *Journal of Positive Behavior Interventions, 13*, 3–16. https://doi.org/10.1177/1098300709359084.

Allison, P. (2011). *Do we really need zero-inflated models?* Statistical Horizons, 7th August https://statisticalhorizons.com/zero-inflated-models.

Amemiya, J., Mortenson, E., & Wang, M. (2020). Minor infractions are not minor: School infractions for minor misconduct may increase adolescents' defiant behavior and contribute to racial disparities in school discipline. *American Psychologist, 5*(1), 23–36. https://doi.org/10.1037/amp0000475.

Artiles, A. J., Kozleski, E. B., Trent, S. C., Osher, D., & Ortiz, A. (2010). Justifying and explaining disproportionality, 1968–2008: A critique of underlying views of culture. *Exceptional Children, 76*(3), 279–299. https://doi.org/10.1177/001440291007600303.

Barger, B., Graybill, E., Roach, A., & Lane, K. (2022). Differential item and test functioning of the SRSS-IE12 across race, ethnicity, gender, and elementary level. *Assessment for Effective Intervention, 47*(2), 79–88. https://doi.org/10.1177/1534508420976830.

Bates, L. A., & Glick, J. E. (2013). Does it matter if teachers and schools match the student? Racial and ethnic disparities in problem behaviors. *Social Science Research, 42*(5), 1180–1190. https://doi.org/10.1016/j.ssresearch.2013.04.005.

Bates, M. E. (2000). Integrating person-centered and variable-centered approaches in the study of developmental courses and transitions in alcohol use: Introduction to the special section. *Alcoholism: Clinical and Experimental Research, 24*(6), 878–881. https://doi.org/10.1111/j.1530-0277.2000.tb02069.x.

Belfield, C., Bowden, A. B., Klapp, A., Levin, H., Shand, R., & Zander, S. (2015). The economic value of social and emotional learning. *Journal of Benefit-Cost Analysis, 6*(3), 508–544. https://doi.org/10.1017/bca.2015.55.

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 national survey. *Journal of School Psychology, 72*, 29–48. https://doi.org/10.1016/j.jsp.2018.12.004.

Blake, J. J., Smith, D. M., Marchbanks, M. P., Seibert, A. L., Wood, S. M., & Kim, E. S. (2016). Does student–teacher racial/ethnic match impact black students' discipline risk? A test of the cultural synchrony hypothesis. In R. Skiba, K. Mediratta, & M. Rausch (Eds.), *Inequality in school discipline* (pp. 79–98). Palgrave Macmillan.

Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces* (Vol. 424). Wiley.

Bronfenbrenner, U. (1994). Ecological models of human development. *Readings on the Development of Children, 2*(1), 37–43.

Burt, K. B., & Roisman, G. I. (2010). Competence and psychopathology: Cascade effects in the NICHD study of early child care and youth development. *Development and Psychopathology, 22*, 557–567. https://doi.org/10.1017/S0954579410000271.

Campbell, S. B., Spieker, S., Burchinal, M., Poe, M. D., & NICHD Early Child Care Research Network. (2006). Trajectories of aggression from toddlerhood to age 9 predict academic and social functioning through age 12. *Journal of Child Psychology and Psychiatry, 47*(8), 791–800. https://doi.org/10.1111/j.1469-7610.2006.01636.x.

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research, 11*, 2079–2107. https://www.jmlr.org/papers/volume11/cawley10a/cawley10a.

Chow, J. C., Ekholm, E., & Coleman, H. (2018). Does oral language underpin the development of later behavior problems? A longitudinal meta-analysis. *School Psychology Quarterly, 33*, 337–349. https://doi.org/10.1037/spq0000255.

Collaborative for Academic, Social, and Emotional Learning (CASEL). (2013). CASEL guide: Effective social and emotional learning programs. http://static.squarespace.com/static/513f79f9e4b05ce7b70e9673/t/526a220de4b00a92c90436ba/1382687245993/2013-casel-guide.pdf.

Collaborative for Academic, Social, and Emotional Learning (CASEL). (2021). SEL research. https://casel.org/research/.

Collins, B. A., O'Connor, E. E., Supplee, L., & Shaw, D. S. (2017). Behavior problems in elementary school among low-income boys: The role of teacher–child relationships. *The Journal of Educational Research, 110*, 72–84. https://doi.org/10.1080/00220671.2015.1039113.

Corcoran, R. P., Cheung, A. C., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review, 25*, 56–72. https://doi.org/10.1016/j.edurev.2017.12.001.

Dishion, T. J., Patterson, G. R., Stoolmiller, M., & Skinner, M. L. (1991). Family, school, and behavioral antecedents to early adolescent involvement with antisocial peers. *Developmental Psychology, 27*, 172–180. https://doi.org/10.1037/0012-1649.27.1.172.

Dodge, K. A., Bai, Y., Ladd, H. F., & Muschkin, C. G. (2017). Impact of North Carolina's early childhood programs and policies on educational outcomes in elementary school. *Child Development, 88*(3), 996–1014. https://doi.org/10.1111/cdev.12645.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., … Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*, 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428.

Duncan, R. J., Schmitt, S. A., Burke, M., & McClelland, M. M. (2018a). Combining a kindergarten readiness summer program with a self-regulation intervention improves school readiness. *Early Childhood Research Quarterly, 42*, 291–300.

Duncan, R. J., Schmitt, S. A., Burke, M., & McClelland, M. M. (2018b). Combining a kindergarten readiness summer program with a self-regulation intervention improves school readiness. *Early Childhood Research Quarterly, 42*, 291–300. https://doi.org/10.1016/j.ecresq.2017.10.012.

Duncan, S. E., & DeAvila, E. A. (1998). *Pre-language assessment scale 2000*. CTB/McGraw-Hill.

Every Student Succeeds Act. (2015). 20 U.S.C. § 6301. https://www.congress.gov/bill/114th-congress/senate-bill/1177.

Feng, Z. D., & McCulloch, C. E. (1996). Using bootstrap likelihood ratios in finite mixture models. *Journal of the Royal Statistical Society: Series B (Methodological), 58*, 609–617. https://doi.org/10.1111/j.2517-6161.1996.tb02104.x.

Figge, C. J., Martinez-Torteya, C., & Weeks, J. E. (2018). Social–ecological predictors of externalizing behavior trajectories in at-risk youth. *Development and Psychopathology, 30*(1), 255–266. https://doi.org/10.1017/S0954579417000608.

Gao, S., Calhoun, V. D., & Sui, J. (2018). Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neuroscience & Therapeutics, 24*, 1037–1052. https://doi.org/10.1111/cns.13048.

Gettinger, M., Kratochwill, T. R., Eubanks, A., Foy, A., & Levin, J. R. (2021). Academic and behavior combined support: Evaluation of an integrated supplemental intervention for early elementary students. *Journal of School Psychology, 89*, 1–19. https://doi.org/10.1016/j.jsp.2021.09.004.

Gillborn, D., Paul, W., & Sean, D. (2018). QuantCrit: Education, policy, 'big Data'and principles for a critical race theory of statistics. *Race Ethnicity and Education, 21*(2), 158–179. https://doi.org/10.1080/13613324.2017.1377417.

Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions. *Yale University Child Study Center, 9*(28), 1–16. https://marylandfamiliesengage.org/wp-content/uploads/2019/07/Preschool-Implicit-Bias-Policy-Brief.pdf.

Girard, L. C., Tremblay, R. E., Nagin, D., & Côté, S. M. (2019). Development of aggression subtypes from childhood to adolescence: A group-based multi-trajectory modelling perspective. *Journal of Abnormal Child Psychology, 47*(5), 825–838. https://doi.org/10.1007/s10802-018-0488-5.

Glass, T. A., Goodman, S. N., Hernán, M. A., & Samet, J. M. (2013). Causal inference in public health. *Annual Review of Public Health, 34*, 61–75. https://doi.org/10.1146/annurev-publhealth-031811-124606.

Gray, S. A., Dueck, K., Rogers, M., & Tannock, R. (2017). Qualitative review synthesis: The relationship between inattention and academic achievement. *Educational Research, 59*, 17–35. https://doi.org/10.1080/00131881.2016.1274235.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system: Manual*. NCS Pearson: American Guidance Service.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*, 127–155. https://doi.org/10.1037/0033-2909.111.1.127.

Horner, R. H., & Sugai, G. (2015). School-wide PBIS: An example of applied behavior analysis implemented at a scale of social importance. *Behavior Analysis in Practice, 8*(1), 80–85.

Jones, S. M., Brown, J. L., & Lawrence Aber, J. (2011). Two-year impacts of a universal school-based social-emotional and literacy intervention: An experiment in translational developmental research. *Child Development, 82*(2), 533–554. https://doi.org/10.1111/j.1467-8624.2010.01560.x.

Jung, T., & Wickrama, K. A. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*(1), 302–317. https://doi.org/10.1111/j.1751-9004.2007.00054.x.

Kagan, S. L., Moore, E. K., & Bredekamp, S. (1995). *Reconsidering children's early development and learning: Toward common views and vocabulary* (Vol. 95, No. 3). National Education Goals Panel.

Kamata, A., Kara, Y., Patarapichayatham, C., & Lan, P. (2018). Evaluation of analysis approaches for latent class analysis with auxiliary linear growth model. *Frontiers in Psychology, 9*, 130. https://doi.org/10.3389/fpsyg.2018.00130.

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology, 3*, 1–27. https://doi.org/10.1146/annurev.clinpsy.3.022806.091432.

Kjellstrand, J. M., Reinke, W. M., & Eddy, J. M. (2018). Children of incarcerated parents: Development of externalizing behaviors across adolescence. *Children and Youth Services Review, 94*, 628–635. https://doi.org/10.1016/j.childyouth.2018.09.003.

Kraemer, H. C., Stice, E., Kazdin, A., Offord, D., & Kupfer, D. (2001). How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *American Journal of Psychiatry, 158*, 848–856. https://doi.org/10.1176/appi.ajp.158.6.848.

Kremer, K. P., Flower, A., Huang, J., & Vaughn, M. G. (2016). Behavior problems and children's academic achievement: A test of growth-curve models with gender and racial differences. *Children and Youth Services Review, 67*, 95–104. https://doi.org/10.1016/j.childyouth.2016.0.

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.

Lam, G. (2014). A theoretical framework of the relation between socioeconomic status and academic achievement of students. *Education, 134*, 326–331. https://doi.org/10.1037/e685842012-030.

Lane, K. L., Oakes, W. P., Common, E. A., Brunsting, N., Zorigian, K., Hicks, T., & Lane, N. A. (2019). A comparison between SRSS-IE and BASC-2 BESS scores at the middle school level. *Behavioral Disorders, 44*(3), 162–174. https://doi.org/10.1177/0198742918794843.

Langer, E. J. (2009). *Counterclockwise: Mindful health and the power of possibility*. Ballantine Books.

Lanza, S. T., & Cooper, B. R. (2016). Latent class analysis for developmental research. *Child Development Perspectives, 10*, 59–64. https://doi.org/10.1111/cdep.12163.

Larsen, R. (2011). Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling: A Multidisciplinary Journal, 18*(4), 649–662. https://doi.org/10.1080/10705511.2011.607721.

Laursen, B., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly, 52*(3), 377–389. https://www.jstor.org/stable/23096200.

Li, M., Johnson, S. B., Musci, R. J., & Riley, A. W. (2017). Perceived neighborhood quality, family processes, and trajectories of child and adolescent externalizing behaviors in the United States. *Social Science & Medicine, 192*, 152–161. https://doi.org/10.1016/j.socscimed.2017.07.027.

van Lier, P. A., Vitaro, F., Barker, E. D., Brendgen, M., Tremblay, R. E., & Boivin, M. (2012). Peer victimization, poor academic achievement, and the link between childhood externalizing and internalizing problems. *Child Development, 83*, 1775–1788. https://doi.org/10.1111/j.1467-8624.2012.01802.x.

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology, 22*, 491–495. https://doi.org/10.1017/S0954579410000222.

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.

Matson, J. L., & Wilkins, J. (2009). Psychometric testing methods for children's social skills. *Research in Developmental Disabilities, 30*(2), 249–274. https://doi.org/10.1016/j.ridd.2008.04.002.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist, 70*, 487–498. https://doi.org/10.1037/a0039400.

McIntosh, K., Horner, R. H., Chard, D. J., Boland, J. B., & Good, R. H. (2006). The use of reading and behavior screening measures to predict non-response to school-wide positive behavior support: A longitudinal analysis. *School Psychology Review, 35*, 275–291.

Metcalfe, L. A., Harvey, E. A., & Laws, H. B. (2013). The longitudinal relation between academic/cognitive skills and externalizing behavior problems in preschool children. *Journal of Educational Psychology, 105*(3), 881–894. https://doi.org/10.1037/a0032624

Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*(4), 674–701. https://ibs.colorado.edu/jessor/psych7536-805/readings/moffitt-1993_674-701.pdf.

Muthén, L. K., & Muthén, B. O. (2018). *Mplus. Statistical analysis with latent variables. User's guide* (7th ed.). Author.

Nagin, D. S., Jones, B. L., Passos, V. L., & Tremblay, R. E. (2018). Group-based multi-trajectory modeling. *Statistical Methods in Medical Research, 27*(7), 2015–2023.

Noltemeyer, A., Palmer, K., James, A. G., & Wiechman, S. (2019). School-wide positive behavioral interventions and supports (SWPBIS): A synthesis of existing research. *International Journal of School & Educational Psychology, 7*(4), 253–262. https://doi.org/10.1080/21683603.2018.1425169.

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science, 26*(5), 617–624. https://doi.org/10.1177/0956797615570365.

Olson, S., Choe, D., & Sameroff, A. (2017). Trajectories of child externalizing problems between ages 3 and 10 years: Contributions of children's early effortful control, theory of mind, and parenting experiences. *Development and Psychopathology, 29*(4), 1333–1351. https://escholarship.org/uc/item/8nc7s7k2.

Owens, A., Reardon, S. F., & Jencks, C. (2016). Income segregation between schools and school districts. *American Educational Research Journal, 53*, 1159–1197. https://doi.org/10.3102/0002831216652722.

Palardy, G., Rumberger, R., & Butler, T. (2015). The effect of high school socioeconomic, racial, and linguistic segregation on academic performance and school behaviors. *Teachers College Record, 117*(12), 1–52.

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*(4), 455–473. https://doi.org/10.1037/edu0000079

Piquero, A. R., Jennings, W. G., & Barnes, J. C. (2012). Violence in criminal careers: A review of the literature from a developmental life-course perspective. *Aggression and Violent Behavior, 17*, 171–179. https://doi.org/10.1016/j.avb.2012.02.008.

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's behavior questionnaire. *Journal of Personality Assessment, 87*(1), 102–112. https://doi.org/10.1207/s15327752jpa8701_09.

Reef, J., Diamantopoulou, S., van Meurs, I., Verhulst, F. C., & van der Ende, J. (2011). Developmental trajectories of child to adolescent externalizing behavior and adult DSM-IV disorder: Results of a 24-year longitudinal study. *Social Psychiatry and Psychiatric Epidemiology, 46*(12), 1233–1241. https://doi.org/10.1007/s00127-010-0297-9.

Reinecke, J., & Seddig, D. (2011). Growth mixture models in longitudinal research. *Advances in Statistical Analysis, 95*(4), 415–434.

Rennie, B., Beebe-Frankenberger, M., & Swanson, H. L. (2014). A longitudinal study of neuropsychological functioning and academic achievement in children with and without signs of attention-deficit/hyperactivity disorder. *Journal of Clinical and Experimental Neuropsychology, 36*, 621–635. https://doi.org/10.1080/13803395.2014.921284.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior assessment system for children* (3rd ed.). Pearson.

Rocque, M., & Paternoster, R. (2011). Understanding the antecedents of the "school-to-jail" link: The relationship between race and school discipline. *The Journal of Criminal Law and Criminology, 101*(2), 633–665.

Sabol, T. J., & Pianta, R. C. (2012). Patterns of school readiness forecast achievement and socioemotional development at the end of elementary school. *Child Development, 83*, 282–299. https://doi.org/10.1111/j.1467-8624.2011.01678.x.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs (report from the governing Board of the American Educational Research Association Grants Program)*. American Educational Research Association. https://www.aera.net/Publications/Books/Estimating-Causal-Effects-Using-Experimental-and-Observational-Designs.

Shiyko, M. P., Li, Y., & Rindskopf, D. (2012). Poisson growth mixture modeling of intensive longitudinal data: An application to smoking cessation behavior. *Structural Equation Modeling: A Multidisciplinary Journal, 19*(1), 65–85. https://doi.org/10.1080/10705511.2012.634722.

Simon, N., & Moore Johnson, S. (2015). Teacher turnover in high-poverty schools: What we know and can do. *Teachers College Record, 117*(3), 1–36.

Skiba, R. J. (2015). Interventions to address racial/ethnic disparities in school discipline: Can systems reform be race-neutral? In R. Bangs, & L. E. Davis (Eds.), *Race and social problems* (pp. 107–124). Springer.

Smolkowski, K., Girvan, E. J., McIntosh, K., Nese, R. N., & Horner, R. H. (2016). Vulnerable decision points for disproportionate office discipline referrals: Comparisons of discipline for African American and white elementary school students. *Behavioral Disorders, 41*, 178–195. https://doi.org/10.2139/ssrn.2753698.

Sullivan, A. L. (2017). Wading through quicksand: Making sense of minority disproportionality in identification of emotional disturbance. *Behavior Disorders, 43*(1), 244–252. https://doi.org/10.1177/0198742917732360.

Sullivan, J. R., Villarreal, V., Flores, E., Gomez, A., & Warren, B. (2021). SSIS performance screening guide as an indicator of behavior and academics: A meta-analysis. *Assessment for Effective Intervention, 46*(3), 228–237. https://doi.org/10.1177/1534508420926584.

Suzuki, S., Morris, S. L., & Johnson, S. K. (2021). Using QuantCrit to advance an anti-racist developmental science: Applications to mixture modeling. *Journal of Adolescent Research, 36*(5), 535–560. https://doi.org/10.1177/07435584211028229.

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L., & Najarian, M. (2018). Early childhood longitudinal study, kindergarten class of 2010-11 (ECLS-K: 2011): User's manual for the ECLS-K: 2011 kindergarten-fourth grade data file and electronic codebook, public version. In *NCES 2018-032*. National Center for Education Statistics. https://doi.org/10.1037/e609692011-001.

Trzesniewski, K. H., Moffitt, T. E., Caspi, A., Taylor, A., & Maughan, B. (2006). Revisiting the association between reading achievement and antisocial behavior: New evidence of an environmental explanation from a twin study. *Child Development, 77*(1), 72–88. https://www.jstor.org/stable/3696691.

Van Dulmen, M. H., Goncy, E. A., Vest, A., & Flannery, D. J. (2009). Group-based trajectory modeling of externalizing behavior problems from childhood through adulthood: Exploring discrepancies in the empirical findings. In J. Savage (Ed.), *The development of persistent criminality* (pp. 288–314). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195310313.003.0014.

Weeks, M., Ploubidis, G. B., Cairney, J., Wild, T. C., Naicker, K., & Colman, I. (2016). Developmental pathways linking childhood and adolescent internalizing, externalizing, academic competence, and adolescent depression. *Journal of Adolescence, 51*, 30–40. https://doi.org/10.1016/j.adolescence.2016.05.009.

Yoon, S., Yoon, D., Wang, X., Tebben, E., Lee, G., & Pei, F. (2017). Co-development of internalizing and externalizing behavior problems during early childhood among child welfare-involved children. *Children and Youth Services Review, 82*, 455–465. https://doi.org/10.1016/j.chiabu.2020.104759.

Zimmermann, C. R. (2018). The penalty of being a young black girl: Kindergarten teachers' perceptions of children's problem behaviors and student–teacher conflict by the intersection of race and gender. *The Journal of Negro Education, 87*(2), 154–168. https://doi.org/10.7709/jnegroeducation.87.2.0154.

Zimmermann, F., Schütte, K., Taskinen, P., & Köller, O. (2013). Reciprocal effects between adolescent externalizing problems and measures of achievement. *Journal of Educational Psychology, 105*(3), 747–761.