HHS Public Access

Author manuscript

Child Youth Serv Rev. Author manuscript; available in PMC 2017 May 18.

Published in final edited form as:

Child Youth Serv Rev. 2016 August ; 67: 95–104. doi:10.1016/j.childyouth.2016.06.003.

Behavior problems and children's academic achievement: A test of growth-curve models with gender and racial differences

Kristen P. Kremer[a,*], **Andrea Flower**[b], **Jin Huang**[a], and **Michael G. Vaughn**[a]

[a]School of Social Work, College for Public Health and Social Justice, Saint Louis University, 3550 Lindell Blvd, St. Louis, MO 63103, United States

[b]Department of Special Education, The University of Texas at Austin, 1912 Speedway, Austin, TX 78712, United States

Abstract

The aim of this study was to examine the longitudinal association between externalizing and internalizing behavior and children's academic achievement, particularly in terms of whether these variables varied as a function of gender and race. Data pertaining to externalizing and internalizing behavior, academic achievement, gender, and race from three waves of the Child Development Supplement of the Panel Study of Income Dynamics ($N = 2028$) were used. Results indicate that behavior problems had a negative relationship with academic performance and some of these associations endured over time. Externalizing behavior impacted reading scores more negatively for females compared to males at baseline, but the impact of externalizing behavior on long-term reading outcomes did not vary by gender. Externalizing behavior impacted reading scores more negatively for Black children than White children at multiple points in time. Differences between males, females, Black, and White children concerning behavior and achievement are explained. Implications, limitations, and ideas for future research are also presented.

Keywords

Achievement; Behavior; Race; Gender; Longitudinal data

1. Introduction

Federal law (No Child Left Behind, NCLB Act of 2001, 2002) mandates academic performance for all children as the top priority for U.S. public schools. NCLB places emphasis on instruction and performance particularly for those with the lowest levels of performance. Although educators work with diligence to achieve at these high levels for all children, often times other factors compromise that progress. For example, some students have extreme academic difficulty that is not easily overcome. Other students have challenging behavior that interferes with teaching and learning. Both of these problems have severe repercussions for the school and life outcomes of these youth (Battin-Pearson et al., 2000; Breslau et al., 2009).

* Corresponding author at: 3550 Lindell Boulevard, St. Louis, MO 63103, United States. kpeter26@slu.edu (K.P. Kremer).

To address these challenges and meet the requirements of NCLB, many schools have adopted multi-tiered systems of support for students who have academic difficulties as well as systems for students who have behavioral difficulties (Doolittle, Horner, Bradley, & Vincent, 2007; Spectrum K12, 2009). These systems are often referred to as a Response to Intervention (RtI) framework within the academic domain and Positive Behavior Interventions and Support (PBIS) within the behavioral domain (Sugai & Horner, 2009). Although use of such systems is frequent, the systems do not tend to be consciously coupled with one another; yet, research (Maguin & Loeber, 1996; Malecki & Elliot, 2002) shows that students with academic problems may also have behavioral problems and that students with behavioral problems may also have academic problems.

The relationship between academic and behavior problems is a long recognized phenomenon (Alexander, Entwisle, & Horsey, 1997; Hinshaw, 1992). A significant amount of research (see Lane, Barton-Arwood, Nelson, & Wehby, 2008; Nelson, Benner, Lane, & Smith, 2004; Reid, Gonzalez, Nordness, Trout, & Epstein, 2004) concerning this relationship comes from the study of students with disabilities such as emotional disturbance (ED) and learning disabilities (LD); yet, as Algozzine, Wang, and Violette (2011) indicate, research on this topic regarding these populations does "little to clarify, confirm, or advance the link between achievement and behavior or the causes for it" (p. 5). In fact, the relationship between achievement and behavior also affects other students, not just those with disabilities: For example as boys from low-income families (Moilanen, Shaw, & Maxwell, 2010) or youth with persistent patterns of externalizing behavior (Vitaro, Brendgen, Larose, & Tremblay, 2005).

Hinshaw (1992) suggested four possibilities for the relationship between academic achievement and behavior including: (a) achievement affects behavior, (b) behavior affects achievement, (c) reciprocal relationships exist between academic and behavioral variables, and (d) some third variable mutually affects behavior and achievement. Although researchers have investigated extensively to understand the relationships between these variables, the relationship remains unclear. Literature has examined the potential impacts of academic achievement on behaviors. In their meta-analysis, Maguin and Loeber (1996) found that poor academic performance appears to be related to frequency, persistence, and seriousness of delinquent activity. A more recent study (Joffe & Black, 2012) revealed that among a sample of 352 secondary school students, those with low academic performance had significantly greater social, emotional, and behavioral difficulties. A variety of research has also suggested that intervention components on the academic domain may have an effect on the behavior domain (Herrenkohl et al., 2001; Maguin & Loeber, 1996).

The current study, however, focuses on predicting academic achievement from behavior problems, the second possibility suggested by Hinshaw (1992). DeLisi and Vaughn's temperament theory can be utilized to understand the process by which behavior predicts academic achievement. According to DeLisi and Vaughn's theory, temperament involves the "stable, largely inborn tendency with which an individual experiences the environment and regulate his or her responses to the environment" (Vaughn, DeLisi, & Matto, 2014, p. 106). Components of temperament include effortful control, the ability to "inhibit a dominant response in favor of performing a subdominant response" (DeLisi & Vaughn, 2014, p. 12),

and negative emotionality, which includes the display of emotions such as frustration, fear, sadness, and discomfort. Inside the classroom, temperament manifests itself as behavior problems and can greatly impact a variety of academic outcomes, including school readiness, elementary school grades, college admission tests, and high school dropout (Duckworth & Allred, 2012; Gumora & Arsenio, 2002).

At school, children have different experiences based on temperament. Research has found, for example, children with low self-control to exhibit poorer work habits than children with higher self-control (Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009). Additional research has found children with lower effortful control to have greater conflict with teachers, while children with higher effortful control have closer relationships with teachers (Rudasill & Rimm-Kaufman, 2009). As a result of these factors, research by Duckworth and Seligman (2005) found self-discipline to be a better predictor of academic performance than IQ. However, the school setting can also enhance academic outcomes for children with difficult temperaments. In particular, research by Rudasill and Rimm-Kaufman (2009) found emotional support from teachers to moderate the relationship between children's temperamental attention and school achievement, while Valiente, Lemery-Chalfant, and Castro (2007) observed school liking to mediate the relationship between children's effortful control and academic competence.

Although the aforementioned research has found temperament and behavior problems to negatively predict academic achievement (e.g., Malecki & Elliot, 2002; Myers, Milne, Baker, & Ginsburg, 1987; Wentzel, 1993), most studies relied on cross-sectional designs and did not take into consideration the nature of behaviors (i.e., externalizing or internalizing behaviors). Few studies investigated the interactions between behaviors and other issues (e.g., race and gender). The present analysis contributes to this topic by investigating the associations of externalizing and internalizing behavior to children's academic achievement in longitudinal data with further examination of whether the association between behavior problems and academic achievement varies by gender and race.

1.1. Externalizing and internalizing behavior related to academic achievement

Externalizing and internalizing behavioral profiles differ greatly. While externalizing behavior is characterized by defiance, disruptiveness, aggressiveness, impulsivity, antisocial behavior, and over-activity, internalizing behavior is marked by withdrawal, dysphoria, and anxiety (Achenbach & Edelbrock, 1978; Hinshaw, 1992). Many researchers have considered the roles of internalizing behavior and externalizing behavior on academic achievement, yet the evidence remains mixed (Masten et al., 2012). Nelson et al. (2004), studying academic achievement of students with emotional/behavioral disorders, found deficits in reading, math, and written language. Math deficits, in particular, appeared to worsen over time—that is, a greater percentage of adolescents with emotional/behavioral disorders performed below average on math measures. Their analyses found no significant differences between males and females on academic achievement measures. Study findings also observed externalizing behaviors to be related to deficits in all three academic areas, while no association was found for internalizing behaviors. On the other hand, a meta-analysis of 26 studies by Riglin,

Petrides, Frederickson, and Rice (2014) found depression, anxiety, and other internalizing behaviors to be associated with increased school failure.

1.2. Race and gender

In addition to the externalizing and internalizing characteristics of behavior, the scholarly conversation about the relationship between academic achievement and behavior problems would be incomplete if other issues were not examined. Race is certainly an issue that should be considered when investigating the relationship between achievement and behavior. Race is a critical factor for several interrelated reasons. First, an achievement gap is widely recognized between Black and White students (Ladson-Billings, 2006; Vanneman, Hamilton, Baldwin Anderson, & Rahman, 2009). Second, Black students continue to be excluded from school through suspension and expulsion at much higher rates than White students (Wallace, Goodkind, Wallace, & Bachman, 2008). Differences in disciplinary practices can contribute to the academic performance gap between Black and White students as when students are suspended from school or the classroom they are provided with fewer opportunities for learning (Gregory, Skiba, & Noguera, 2010). Unfortunately, suspension is linked to a number of problems including dropout, which has also been linked to academic performance (Allensworth & Easton, 2007) and Black youth dropout of school at a greater rate than White youth (Chapman, Laird, Ifill, & Kewal Ramani, 2010). Racial differences in suspension rates can be directly linked to differences in problem behaviors. While many researchers suggest racial disparities in suspension rates result from racial discrimination, recent findings by Wright, Morgan, Coyne, Beaver, and Barnes (2014) indicates that racial differences in suspension rates can be explained by prior behavior problems. Finally, Black students are consistently over-identified for special education eligibility particularly in the categories of Emotional Disturbance and Intellectual Disability (Hosp & Reschly, 2004), which are disability categories with academic and behavioral risk profiles. In addition to the aforementioned link between academic achievement and behavior, behavior problems may disproportionately impact academic outcomes for Black students. In particular, Rabiner, Murray, Schmid, and Malone (2004) found challenging behavior in the form of inattention to be rated higher among White students, yet inattention demonstrated a negative association with academic achievement for Black students but not White students.

Research has also shown an achievement gap between males and females. Historically, achievement and academic performance has favored males over females particularly in math and science (Weaver-Hightower, 2003). Although, a number of studies suggest that females have historically outperformed boys in the area of literacy (Gambell & Hunter, 1999; Ready, LoGerfo, Burkam, & Lee, 2005) and may begin school with greater literacy skills than boys (Ready et al., 2005). One possible explanation for this performance difference is classroom behavior. For example, using data from the Early Childhood Longitudinal Study-Kindergarten Cohort, Ready et al. (2005) found that approaches to learning and children's problematic behaviors explained 70% and 15% of the variance in literacy performance, respectively, with girls receiving better ratings in these areas than boys. Additionally, boys have significantly higher rates of grade retention (Meisels & Liaw, 1993; Ready et al., 2005), suspensions, and expulsions than girls (Gregory et al., 2010). Males are also overrepresented in disability groups such as emotional/behavioral disorders (Coutinho & Oswald, 2005).

While a number of the issues highlighted here have been widely studied and attended to, only a few of these variables have been studied simultaneously. For example, the effect of externalizing and internalizing behavior on academic achievement has been somewhat minimally examined (e.g., Arnold, 1997; Lane et al., 2008; Nelson et al., 2004; Nelson, Benner, Neill, & Stage, 2006; Ready et al., 2005), but many of these studies (Lane et al.; Nelson et al., 2004; Nelson et al., 2006) as stated earlier, concern populations with disabilities defined by academic and behavioral difficulties. Furthermore, studies investigating such relationships using samples of students who do not have disabilities include information about students given by teacher rather than parent rating scales or direct observation. Finally, most studies are cross-sectional rather than longitudinal and those that are longitudinal (e.g., Ready et al., 2005) do not consider all of these variables (i.e., academic and type of behavior problems with race and gender).

1.3. Present study aim

The transactional and progressive associations between behavior problems (externalizing and internalizing) and academic achievement can only be detected in longitudinal studies. Nonetheless, there remain surprisingly few longitudinal studies that have been conducted to examine these links (Bub, McCartney, & Willett, 2007; Chen & Li, 1997; McGee, Feehan, Williams, & Anderson, 1992); and even fewer that appear to use large national datasets (Bub et al., 2007). This study aims to fill this gap by examining the unfolding of the relationship between externalizing and internalizing behaviors and academic achievement from age 3 to 17 in a nationally representative sample of children. In addition, this analysis allows for examination of the heterogeneity in the association between externalizing and internalizing behaviors and academic achievement by gender and race. Focusing on the second mechanism proposed by Hinshaw (1992), two main hypotheses were tested in this study. First, it is hypothesized that behavior problems will be associated with decreased academic performance contemporaneously and will be negatively associated with the change of academic performance over time. Second, the association between behavior problems and academic performance will differ by gender or race.

2. Method

2.1. Data and sample selection

The study objectives were examined using data from the three waves of the Child Development Supplement (CDS) of the Panel Study of Income Dynamics (PSID). The PSID is a longitudinal survey that collected demographic information and socioeconomic characteristics from a nationally representative sample of individuals and their families annually between 1968 and 1997 and biennially thereafter. Beginning in 1997, the PSID supplemented its core data with additional information from a group of children 0–12 years old ($N = 3563$) in the Child Development Supplement (CDS). The same children were interviewed three times in 1997, 2002, and 2007, respectively, if they were still younger than age 18 at the time of each interview. The recruiting, eligibility, and attrition of the PSID-CDS have been described elsewhere (Institute for Social Research, 2010). The CDS included measures of a broad array of child developmental outcomes, such as physical health and disability, emotional well-being, cognitive and academic achievement, and social

relationships with household members and peers. The CDS collected the information on behavior problems for children older than three in the first wave and conducted standardized achievement tests on children in all three waves.

To examine the relationship between behavior problems and the long-term performance of academic achievement, Black and White children who had at least one valid measure on standardized achievement tests among three waves and had valid information on behavior problems measured in the first wave ($N = 2143$) were included in the analysis. The sample participants with only one valid measure among three repeated standardized achievement tests were included because they still contribute to the estimation of the development of academic achievement at a specific time point. In addition, the few children with missing values on variables listed in Table 1 were excluded ($n = 115$). The excluded children had missing information mainly on the variables of birth weight ($n = 43$), mother's education ($n = 53$), mother's age ($n = 21$), and mother's employment status ($n = 21$); the final sample size was 2028. Since only about 5% of children had missing information, we used listwise deletion in main analyses reported below.

2.2. Outcome variables

The outcome variables were children's scores on three subtests of the Woodcock-Johnson Revised (WJ-R) Tests of Achievement (Woodcock & Johnson, 1989) across three waves. As a standardized measure of child's academic skill, the Woodcock-Johnson Revised (WJ-R) Tests of Achievement have been widely used and have demonstrated excellent reliability and validity. The PSID-CDS administered the Letter-Word Identification test (LW subtest) and the Applied Problems test (AP subtest) on children who were three years or older, and the Passage Comprehension test (PC subtest) on those who were six years or older in all three waves. The LW and PC are two subtests on children's reading ability, and the AP concerns children's math ability. The three WJ-R subtests are individually administered scales. The LW subtest measures children's reading skills particularly their ability to name letters and sounds as well as to decode words. The PC subtest measures children's comprehension and vocabulary skills using multiple-choice questions. The AP subtest measures children's ability to solve math problems for applied purposes (e.g., determine whether there is enough money available to purchase items shown on the test page given the coins shown.) The WJ-R Test of Achievement standardizes the raw scores of three tests to a 0–200 continuous variable, respectively; all the WJ-R scores presented in the study thus are standardized ones.

2.3. Independent variables

The analysis included four major independent variables: children's age (3–17), gender (1 = male and 0 = female), race (1 = Black and 0 = White), and behavior problems. Since the PSID was initiated in 1968, the sample included relatively fewer households with racial background different from Black and White, and those participants were not included in the study sample. The PSID-CDS included a 32-item Behavior Problem Index (BPI) developed by Peterson and Zill (1986) to measure the incidence and severity of child behavior problems (externalizing and internalizing). It was based on primary caregiver's responses for children three years and older as to whether a set of problem behaviors was often ("3"), sometimes ("2"), or never ("1") true. The index was divided into two subscales, one for

externalizing (α = 0.86) or aggressive behavior (16 items), and the other one for internalizing (α = 0.81), withdrawn or sad behavior (13 items). The externalizing subscale included items such as the child "is impulsive and acts without thinking" and "argues too much". The internalizing subscale had items such as the child "is withdrawn and has trouble getting involved with other children" and "is unhappy, sad or depressed". Continuous summary scores from these items were used to assess externalizing and internalizing behaviors. Except for children's age, all independent variables are measured in the first wave of the Panel Study of Income Dynamics (2012).

2.4. Covariates

Adjustments were made for three groups of covariates measured in 1997, but for presentation purposes only the coefficients for the key explanatory variables (age, gender, race, and behavior problems) were reported. The first group consisted of children's health and disability characteristics, including preterm birth (1 = gestational age less than 37 weeks and 0 = others), low birth weight (1 = birth weight less than 2500 g and 0 = others), neonatal intensive care at birth (1 = yes and 0 = no), and physical/mental limitations on childhood or school activities (1 = yes and 0 = no). As shown in the literature (e.g., Lane et al., 2008; Nelson et al., 2004; Reid et al., 2004), disability status of children is an important predictor for both academic achievement and behavior problems.

The second group of covariates consisted of mothers' characteristics, including age, number of schooling years (1–17 years), employment status (1 = yes and 0 = no), and mothers' parenting practices. Three indicators of parenting practices created by the PSID-CDS were controlled for: parental warmth, emotional support, and cognitive stimulation, because parenting practices are likely to associate with both children's academic achievement and behavior problems (e.g., Fan & Chen, 2001; Jeynes, 2007; Shumow, Vandell, & Posner, 1998; Stormshak, Bierman, McMahon, & Lengua, 2000). Ranging from 1–5, parental warmth was a standardized scale reported by parents and measuring the warmth of the relationship between mothers and children, including the frequency of showing physical affection, appreciation, and so on. Emotional support and cognitive stimulation were based on observed interactions between mothers and children in home environment by the well-trained interviewers employed by the PSID-CDS. Interviewers were extensively trained in techniques and procedures of general interviews and specific data collection, including unique protocols to conducting observations on emotional support and cognitive simulation. The interviewer recorded observed home environment and interactions between the child and the primary caregivers after the in-home interview. Emotional support ranged from 2–14, and was summarized from items observed by interviewers, such as "mother caressed, kissed, or hugged child at least once" and "mother conversed with child at least twice." Cognitive stimulation ranged from 2–14, and included items such as "how many books child has read" and "mother provided toys or interesting activities." Having demonstrated their validity and reliability in previous surveys (e.g., the National Longitudinal Survey of Youth; Panel Study of Income Dynamics, 2012), all of these three scales have the value of Cronbach's alpha greater than 0.82 in the data. The correlation among three parenting indicators is lower than 0.32.

Finally, adjustments for household characteristics: household size, number of children in households, food stamp program participation (1 = yes and 0 = no), Aid to Families with Dependent Children (AFDC) or Temporary Assistance for Needy Families (TANF) participation (1 = yes and 0 = no), homeownership (1 = yes and 0 = no), and log-transformed household income. The analyses also controlled for dichotomous indicators of residence states of the PSID-CDS children, which may serve to impart regional socialization differences on study children.

2.5. Data analyses

In order to examine the long-term performance of children's academic achievement, a growth curve analysis for each test score in the multilevel modeling context (Rabe-Hesketh & Skronda, 2012) was conducted, and the association between behavior problems measured in 1997 and the changes of children's academic achievement over time is tested. The model specification is listed below:

$$Y_{ti}=\beta_0+\beta_1*A_{ti}+\beta_2*A^2{}_{ti}+\beta_3*G_i+\beta_4*R_i+\beta_5*B_i+\beta_6*(A_{ti}*B_i)+\beta_7*X_i+u_{0i}+u_{1i}*A_{ti}+\varepsilon_{ti} \quad (1)$$

where Y_{ti} indicates test score of a child *i* at wave *t*; A_{ti} denotes the age of child *i* at wave *t*; G_i and R_i refer to gender and race of child *i*; B_i indicates behavior problems (either externalizing or internalizing subscale) measured in the first wave; and X_i is a vector of control variables, including all three groups of covariates discussed above. We only included measures of behavior problems in the first wave rather than changes of behavior problems over time because the study focused on the potential mechanism that behavior affects achievement (Hinshaw, 1992). The relationships between changes of behavior problems and academic achievement may instead reflect reciprocal relationships that exist between academic and behavioral variables.

The study centered child's age to the first year when they were allowed to take specific tests (i.e., age 3 for the LW and AP tests and age 6 for the PC test). A quadratic term of children's age (A^2_{ti}) was used to model the potential nonlinearity of the growth in academic achievement over time. A cubic term of children's age in a sensitivity test was also added, which generated consistent results with those reported below. As indicated by two random parts (u_{0i} and $u_{1i} * A_{ti}$), the intercept and the regression coefficient on children's age were allowed to vary by child in the growth curve model. However, the regression coefficient of the quadratic term of child's age was not set as a random one. The analysis showed the variance of this coefficient is very small, if it was defined as a random coefficient. To test the first hypothesis, the parameters of interest in this equation are β_5 and β_6— β_5 indicates the association between children's behavior problems and their academic achievement at baseline, and β_6 shows the association between behavior problems and the performance of academic achievement over time.

To examine whether the association between behavior problems and academic achievement varies by gender, the two interaction terms were added to Eq. (1)—an interaction between behavior problems and gender and a three-way interaction among behavior problems, gender, and age. The former one shows the difference in the relationship between behavior problems and academic achievement at baseline by gender, and the latter suggests whether the association between behavior problems and the changing rate of academic achievement differs by gender. The same strategy—adding an interaction between behavior problems and race and a three-way interaction among behavior problems, race, and age into Eq. (1)—was used to investigate the potential heterogeneity in the association between behavior problems and academic achievement by race.

In order to decrease complexity, results for control variables are not shown in regression models (Tables 3–5). For all statistical analyses, weighted estimates to account for the oversampling of minority children and data attrition and standard errors were computed using Stata 12.1SE (StataCorp, 2011). This approach implements a Taylor series linearization to adjust standard errors of estimates for complex survey sampling design effects including clustered data (StataCorp, 2011). The strategies discussed above added measures of externalizing and internalizing behavior problems separately into analyses to avoid the potential multicollinearity between the two; however, this approach limits the opportunity to explore the interaction between externalizing or internalizing behavior problems on academic achievement. We thus included both measures in the same analyses in robustness tests. Results were generally consistent with those reported below, but with smaller and less significant regression coefficients for externalizing and internalizing problems.

3. Results

3.1. Descriptive analysis

Table 1 reports the distributions for the outcome and independent variables and demographic characteristics of the analytic sample. The mean standardized scores of the AP, LW, and PC tests were 107.3 (SD = 16.6), 105.5 (SD = 18.4), and 105.2 (SD = 17.7), respectively. The PSID-CDS children, on average, were 8.4 years old in 1997. The age range of these children was from 3–12 in 1997, with a sample size from 150 to 200 for each age. While not reported in Table 1, the age range of these children was about 8–17 in 2002 when the second wave of the PSID-CDS was conducted. The sample size for each age in the second wave was from 130 to 190. Since the PSID-CDS only collected information for children younger than 18, the age range was approximately 13–17 in 2007 (i.e., the third wave), and the sample size for children aged 17 is reduced to about 100 because some children aged out from the PSID-CDS. Half of the children were male, and nearly 80% of subjects were White. Mean summary scores for the externalizing and internalizing subscales were 23.1 (SD = 5.7) and 16.0 (SD = 3.6) respectively. The mean age of their mothers was 36.0 (SD = 7.7) in 1997, and the mean schooling years for mothers was 13.1 (SD = 2.7). More than 60% of mothers were employed in 1997. On average, children lived in a household with four members (including two children) and reported a mean income of about $57,000.

Fig. 1 presents the average scores of the three tests by children's age. The range of the y-axis is nearly two standard deviations around the mean test score. For the AP, LW, and PC tests, the mean scores at different ages were connected using the dashed line, the solid line, and the dotted line, respectively. The figure shows that, overall, the mean score was about 5 points above or below 100. There was a slight upward trend in early childhood, and a slight downward trend was noticed in late childhood.

3.2. Bivariate analysis

Reporting average test scores over three waves, Table 2 confirms that children's academic achievement and behavior problems vary by gender and race. Female children had higher mean scores on the LW and PC tests of about 2.5 points ($p < 0.001$). Conversely, male children had greater mean scores on the AP test (108.9 vs. 105.6, $p < 0.001$) and the externalizing subscale (23.6 vs. 22.5, $p < 0.001$). Males and females do not differ significantly with respect to the internalizing subscale.

With respect to race, Black children had test scores nearly 15 points lower than White children on all three tests of academic performance ($p < 0.001$ level), and their mean score on the externalizing subscale was 0.7 points higher than that of White children (23.6 vs. 22.9, $p < 0.001$).

3.3. Associations between behavior problems and academic achievement

Table 3 reports results of an Eq. (1) using the externalizing or internalizing subscale to predict all three test scores while adjusting for previously described control variables. First, except for the PC score, we found statistically significant and positive regression coefficients on the age variable and negative coefficients on the age-squared variable, which suggests that there is a curvilinear relationship between age and children's achievement test scores. The positive marginal effects of age on test scores decrease when children get older. Second, controlling for behavior problems and all other variables in the model, male children had LW and PC scores about 1.5 points lower than female children and had AP scores three points higher than female children at baseline. The Black-White achievement gap was approximately seven points on two reading tests and nearly 11 points on the AP test at baseline.

Children's externalizing behavior problems were inversely associated with all three test scores at baseline. Specifically, a one-point increase in the externalizing subscale reduced children's AP score by 0.35 points at age 3 (95% CI: −0.53, −0.17; $p < 0.001$), decreased the LW score by 0.22 points at age 3 (95% CI: −0.40, −0.03; $p < 0.05$), and lowered the PC score by 0.22 points at age 6 (95% CI: −0.39, −0.05; $p < 0.05$). While the externalizing subscale did not affect the change of the LW and PC scores over time, its interaction with children's age was significant in the model predicting the AP score ($b = 0.02$, 95% CI: 0.00, 0.04; $p < 0.05$). It indicates the negative associations between externalizing behaviors and the LW and PC scores remain consistent across ages, while the association between externalizing behaviors and the AP score weakens over time. Fig. 2 indicates that a typical child (see the definition of the typical child in Fig. 2) with the externalizing subscale at the third quartile point (= 26) has an AP score two points lower than a typical child with the

externalizing subscale at the first quartile point (= 19) at baseline, but the score difference reduces over time and disappears at about age 15.

Children's internalizing behavior problems were negatively associated with the AP and PC scores but not with the LW score at baseline. A one-point increase in the internalizing subscale reduced children's AP score by 0.50 points (95% CI: −0.80, −0.20; p < 0.000) and the PC score by 0.35 points (95% CI: −0.64, −0.06; p < 0.01), respectively.

3.4. Associations between behavior problems and academic achievement by gender

Results of the associations between behavior problems and academic achievement by gender are reported in Table 4. The interaction term between the externalizing subscale and gender is significant in the models using the LW score (b = 0.25, 95% CI: 0.02, 0.48, $p < 0.05$), suggesting that externalizing behavior programs have greater associations with decreased academic performance for female children at baseline. None of the three-way interaction terms among gender, age, and behavior problems were statistically significant in the models. Based on the results of the second column in Table 4, Fig. 3 predicts the LW scores over time for four typical cases with different genders and different levels of externalizing. The slopes, indicating the changing rates of reading achievement, are almost the same for the four predicted lines. At baseline, the difference in the LW test score was nearly 3 points for female children on the externalizing subscale at the first quartile point (= 19) and the third quartile point (= 26), but was only 1.2 points for male children.

3.5. Associations between behavior problems and academic achievement by race

Results on the associations between behavior problems and academic achievement by race are presented in Table 5. The interaction terms between behavior problems and race were not statistically significant in these models; the associations between behavior problems and academic performance do not vary by race at baseline. However, different from the results on gender, the three-way interaction term among behavior problems (either externalizing or internalizing subscale), age, and race were negatively associated with children's reading test scores (i.e., LW and PC scores). A one-point increase in the externalizing or internalizing subscale at baseline decreases Black children's reading score about 0.02 points more than White children every year. That is, in addition to their associations with academic performance at baseline, behavior problems are negatively correlated with reading scores over time for Black children. Based on the results shown in the fifth column in Table 5, Fig. 4 predicts the LW scores over time for two typical cases with the internalizing subscale at the mean level (= 15). The Black-White test gap was 5.1 points at age 3, but reached 11.2 points at age 17.

4. Discussion

The purpose of this study was to examine the relationship between externalizing and internalizing behaviors and the long-term trajectory of academic achievement in a nationally representative sample of children. The heterogeneity in the relationship between externalizing and internalizing behaviors and academic achievement by gender and race was examined. It was hypothesized that behavior problems are associated with decreased

academic performance at baseline and that negative association continues over time (hypothesis 1). Furthermore, it was hypothesized that the association between behavior problems and academic performance differ by gender and race (hypothesis 2). Findings suggested that both hypotheses were at least partially correct.

At baseline, behavior problems did appear to have a negative relationship with academic performance wherein externalizing behavior impacted all three academic subtests (i.e., LW, AP, and PC) and internalizing behavior impacted PC and AP. Interestingly, the association of externalizing behavior with the AP score faded over time. However, the effect of externalizing behavior remained for LW and PC over time (hypothesis 1). Males also had higher externalizing scores than females and females performed higher than males on reading measures, which is consistent with previous research (Ready et al., 2005). Although, interestingly, even though males scored higher on externalizing problems and, overall, lower on reading measures, externalizing behaviors appeared to have greater negative impacts on female children's reading achievement in baseline than male children's achievement (hypothesis 2). However, the long-term trajectories did not appear to be affected.

Our findings with regard to math performance over time appear to contradict previous findings (Nelson et al., 2004). The difference in results between our study and that of Nelson et al. could be due to a variety of reasons. First, Nelson et al. worked with a sample with 155 youth with emotional/behavioral disorders and we based our study on a large national sample. Second, the data that we used were from a longitudinal data set rather than obtained through a cross-sectional design. Finally, Nelson et al. included multiple measures of math achievement from the Woodcock-Johnson III: math calculation, math fluency, and applied problems; whereas, the PSID dataset only included applied problems (AP).

Overall, our results agreed with Hinshaw's (1992) theory which suggested that achievement and behavior are related. These results suggest that there is an inverse relationship between achievement and behavior and that this relationship has lasting effects over time, particularly for reading scores.

Our findings also corroborated results from other research demonstrating an achievement gap between Black and White students, with Black children performing lower than White children on the academic measures and significantly higher on the externalizing behavior measure (e.g., Rabiner et al., 2004). Most significantly, the results of the analyses suggested that behavior problems had a greater effect on Black children's reading achievement as the children aged than on White children's reading achievement as they aged (hypothesis 2). Using results in Table 5 to predict LW scores by age, race, and internalizing problems, the differences found between Black and White students at baseline were approximately one-third of a standard deviation on the LW subtest, but 14 years later that gap had widened to nearly three-quarters of a standard deviation. While Black and White children's scores were just above average on LW (M = 100, SD = 15) at age 3, by age 17 White children's scores remained near average and Black children's scores dipped into the low average range. Even using parent-rating scales of children's behavior, behavioral problems continued to affect students' achievement. Specifically, behavioral problems appeared to affect females' and Black students' reading achievement.

These analyses have benefits over some of the previous research conducted concerning these topics: (1) the data set analyzed here was a longitudinal data set and the findings are based on performance over time rather than on a cross-sectional dataset, and (2) this analysis also simultaneously considered the relationship between externalizing behavior, internalizing behavior, race and gender, which few previous studies considered with a large sample.

4.1. Implications

A few important implications result from this study. First, even though behavior problems are concerns by themselves, the effect of behavior on reading skills is an especially critical finding particularly given research (Allensworth & Easton, 2007), which suggests that course failure in English is a predictor for later school dropout. Our study revealed that the impact of behavior problems remains long-term for LW and PC, two subtests on children's reading ability. This is an especially critical finding given that coursework becomes more reading intensive as students progress through school and may have critical implications for youth with externalizing behavior as they reach high school, particularly as these students may be most at risk for school dropout given that course failure and behavior problems uniquely contribute to dropout (Allensworth & Easton, 2007). Practitioners need to consider how the interaction between externalizing behavior and reading difficulty affects student performance in classes that require significant reading and comprehension of text. In our study, externalizing behavior did not appear to affect math achievement as greatly as it did reading performance. This is an important implication for school-based screening of students with academic and behavior difficulties: measures of reading achievement and behavior might be useful in determining which students may need greater support. Although we did not find significant effects of behavior on math performance, previous research has found that math is greatly affected (Nelson et al., 2004) particularly across time; thus, it remains an important consideration especially in light of its status as a predictor for school dropout by middle school and high school.

Our findings regarding the association between behavior and achievement suggest that school professionals should consider providing intervention in both domains to the same students; a reasonable way to do this may be to couple Response to Intervention and Positive Behavioral Interventions and Supports in schools. Such multi-tiered systems of support are also needed early, before children have really had the opportunity to begin the cycle of failure. This means that educators must remain vigilant in identifying students at-risk in these areas, perhaps by frequent screening. Given the findings that behavior problems may affect Black children's reading achievement more profoundly and the conclusions of previous research that indicate overrepresentation issues pertaining to Black students in special education (Hosp & Reschly, 2004), efficient and effective early intervention appears critical for these students.

Furthermore, in line with DeLisi and Vaughn's temperament theory, changes within the classroom can enhance academic outcomes for children with difficult temperaments. Considering Rudasill et al.'s (2010) findings that teacher's emotional support moderates the relationship between temperament and academic outcomes, training can be provided to foster teachers' emotional support. In particular, Strengthening Emotional Support Services

is a curriculum designed to equip teachers with behavioral management strategies and minimize classroom disruption, and has been found to increase academic engagement for students with behavioral and emotional disorders (Sawka, McCurdy, & Mannella, 2002). Additionally, given aforementioned researched by Rudasill and Rimm-Kaufman (2009), increased teacher training on relationship-building with children of all temperaments may also enhance academic outcomes.

4.2. Limitations

While this study offers improvements over research conducted with cross-sectional research designs, limitations are still present. First, this study was based on the PSID dataset, a large national, longitudinal dataset. The analyses presented here rely on the availability and quality of the data contained in PSID. For example, only three subsets of the WJ-R Tests of Achievement are available. The data collected for PSID were initially collected annually and then biannually after 1997. This means that data were not available yearly on outcomes after 1997; however, the growth curve modeling employed for analysis in this study does not require continuous (i.e., annual) data. In addition, the current study only uses behavior problems measured at Wave 1 to predict children's long-term academic achievement and does not consider the changes of behavior problems in analyses, which may affect children's academic achievement. Moreover, although various confounding variables were controlled for, there may be other child, parental, or school context characteristics that could possibly threaten these findings; for example, teacher-student relationships, quality of classroom management, and availability of resources at the school may be related to achievement as well. Finally, the reliance on parent-reported behavior may bias findings. While parent report of internalizing behavior has been found to be more highly associated with observed behavior than teacher reports, the opposite has been found for externalizing behavior (Hinshaw, Han, Erhardt, & Huber, 1992). Additionally, research has found parents to report greater problems with externalizing behavior than teachers (Verhulst & Akkerhuis, 1989; Stanger & Lewis, 1993). Bias could be reduced by utilizing reports from both parents and teachers, as has been suggested by Verhulst, Koot, and Van der Ende (1994) who found the use of parent and teacher reports together, rather than one or the other, to increase predictive power.

4.3. Future research

The finding that behavior problems appear to more negatively affect Black children's reading thus widening the achievement gap gives rise to questions about how such a gap might be narrowed or rather how it can be ensured that all children perform at the height of their ability—that is, as a group, closer to the average on such norm-referenced or standardized tests. Future research might investigate whether all children with behavior problems should be provided with additional reading assistance or at the very least close progress monitoring in the area of reading. Moreover, future research might address how Response to Intervention and Positive Behavioral Interventions and Supports might be more closely aligned to be sure that children with behavioral problems are monitored for academic needs and that children with academic needs are monitored for behavioral problems. Finally, future research should continue to investigate the relationship between behavior problems and academic achievement with the realization that these issues might be related bi-

directionally—that is, academic or behavior problems could be driving the other. Future researchers might work to distinguish directionality between the achievement and behavior variables using longitudinal data. This research is likely to have serious implications for how practitioners identify and intervene with at risk learners in today's schools.

Acknowledgements

The authors are grateful for support from the Meadows Center for Preventing Educational Risk, the Institute on Educational Sciences grants (R324A100022 & R324B080008) and from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (P50 HD052117). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute Of Child Health and Human Development or the National Institutes of Health.

References

Achenbach TM, Edelbrock CS. The classification of child psychopathology: A review and analysis of empirical efforts. Psychological Bulletin. 1978; 85:1275–1301. http://dx.doi.org/10.1037/00332909.85.6.1275. [PubMed: 366649]

Algozzine B, Wang C, Violette AS. Reexamining the relationship between academic achievement and social behavior. Journal of Positive Behavior Interventions. 2011; 13:3–16. http://dx.doi.org/10.1177/1098300709359084.

Allensworth, EM., Easton, JQ. What matters to staying on-track and graduating from Chicago public high schools. Consortium on Chicago School Research; Chicago: 2007.

Alexander KL, Entwisle DR, Horsey CS. From first grade forward: Early foundations of high school dropout. Sociology of Education. 1997; 70:87–107.

Arnold DH. Co-occurrence of externalizing behavior problems and emergent academic difficulties in young high-risk boys: A preliminary evaluation of patterns and mechanisms. Journal of Applied Developmental Psychology. 1997; 18:317–330. http://dx.doi.org/10.1016/S0193-3973(97)80003-2.

Battin-Pearson S, Newcomb MD, Abbott RD, Hill KG, Catalano RF, Hawkins JD. Predictors of early high school dropout: A test of five theories. Journal of Educational Psychology. 2000; 92:568–583. http://dx.doi.org/10.1037/0022-0663.92.3.568.

Breslau J, Miller E, Breslau N, Bohnert K, Lucia V, Schweitzer J. The impact of early behavior disturbances on academic achievement in high school. Pediatrics. 2009; 123:1472–1476. http://dx.doi.org/10.1542/peds.2008-1406. [PubMed: 19482756]

Bub KL, McCartney K, Willett J. Behavior problem trajectories and first-grade cognitive ability and achievement skills: A latent growth curve analysis. Journal of Educational Psychology. 2007; 99:653–670. http://dx.doi.org/10.1037/0022-0663.99.3.653.

Chapman, C., Laird, J., Ifill, N., Kewal Ramani, A. Trends in high school dropout and completion rates in the United States: 1972–2008 (NCES 2011-012). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; Washington, DC: 2010.

Chen X, R. K. H, Li D. Relation between academic achievement and social adjustment: Evidence from Chinese children. Developmental Psychology. 1997; 33:518–525. http://dx.doi.org/10.1037/0012-1649.33.3.518. [PubMed: 9149930]

Coutinho MJ, Oswald DP. State variation in gender disproportionality in special education: Findings and recommendations. Remedial and Special Education. 2005; 26:7–15. http://dx.doi.org/10.1177/07419325050260010201.

DeLisi M, Vaughn M. Foundation for a temperament-based theory of antisocial behavior and criminal justice system involvement. Journal of Criminal Justice. 2014; 42:10–25.

Doolittle JH, Horner RH, Bradley S, Vincent C. Importance of student social behavior in the mission, statements, personnel preparation standards, and innovation efforts of state departments of education. The Journal of Special Education. 2007; 40:239–245. http://dx.doi.org/10.1177/00224669070400040501.

Duckworth, AL., Allred, KM. Temperament in the classroom. In: Zenter, M., Shiner, RL., editors. Handbook of temperament. Guilford Press; New York: 2012. p. 627-644.

Duckworth AL, Seligman ME. Self-discipline outdoes IQ in predicting academic performance of adolescents. Psychological Science. 2005; 16(12):939–944. [PubMed: 16313657]

Fan X, Chen M. Parental involvement and students' academic achievement: A meta-analysis. Educational Psychology Review. 2001; 13(1):1–22.

Gambell TJ, Hunter DM. Rethinking gender differences in literacy. Canadian Journal of Education. 1999; 24:1–16.

Gregory AG, Skiba RJ, Noguera PA. The achievement gap and the discipline gap: Two sides of the same coin. Educational Researcher. 2010; 39:59–68. http://dx.doi.org/10.3102/0013189X09357621.

Gumora G, Arsenio WF. Emotionality, emotion regulation, and school performance in middle school children. Journal of School Psychology. 2002; 40(5):395–413.

Herrenkohl TI, Guo J, Kosterman R, Hawkins JD, Catalano RF, Smith BH. Early adolescent predictors of youth violence as mediators of childhood risks. The Journal of Early Adolescence. 2001; 21:447–469. http://dx.doi.org/10.1177/0272431601021004004.

Hinshaw SP. Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. Psychological Bulletin. 1992; 111:127–155. [PubMed: 1539086]

Hinshaw SP, Han SS, Erhardt D, Huber A. Internalizing and externalizing behavior problems in preschool children: Correspondence among parent and teacher ratings and behavior observations. Journal of Clinical Child Psychology. 1992; 21(2):143–150.

Hosp JL, Reschly DJ. Disproportionate representation of minority students in special education: Academic, demographic, and economic predictors. Exceptional Children. 2004; 70:185–199.

Jeynes WH. The relationship between parental involvement and urban secondary school student academic achievement: A meta-analysis. Urban Education. 2007; 42(1):82–110. http://dx.doi.org/10.1177/0042085906293818.

Joffe VL, Black E. Social, emotional, and behavioral functioning of secondary students with low academic and language performance: Perspectives from students, teachers, and parents. Language, Speech, and Hearing Services in Schools. 2012; 43:461–473. http://dx.doi.org/10.1044/0161-1461(2012/11-0088).

Ladson-Billings G. From the achievement gap to the education debt: Understanding achievement in U.S. schools. Educational Researcher. 2006; 35(7):3–12. http://dx.doi.org/10.3102/0013189X035007003.

Lane KL, Barton-Arwood SM, Nelson JR, Wehby J. Academic performance of students with emotional and behavioral disorders served in a self-contained setting. Journal of Behavioral Education. 2008; 17:43–62.

Maguin E, Loeber R. School performance and delinquency. Journal of Crime and Justice. 1996; 20:145–264.

Malecki CK, Elliot SN. Children's social behaviors as predictors of academic achievement: A longitudinal analysis. School Psychology Quarterly. 2002; 17:1–23. http://dx.doi.org/10.1521/scpq.17.1.1.19902.

Masten AS, Roisman GI, Long JD, Burt KB, Obradović J, Riley JR, Tellegen A. Developmental cascades: Linking academic achievement and externalizing and internalizing symptoms over 20 years. Developmental Psychology. 2012; 41(5):733–746. http://dx.doi.org/10.1037/0012-1649.41.5.733.

McGee R, Feehan M, Williams S, Anderson J. DSM-III disorders from age 11 to age 15 years. Journal of the American Academy of Child and Adolescent Psychiatry. 1992; 31:50–59. http://dx.doi.org/10.1097/00004583-199201000-00009. [PubMed: 1537781]

Meisels SJ, Liaw F. Failure in grade: Do retained students catch up? The Journal of Educational Research. 1993; 87:69–77. http://dx.doi.org/10.1080/00220671.1993.9941169.

Moilanen KL, Shaw DS, Maxwell KL. Developmental cascades: Externalizing, internalizing, and academic competence from middle childhood to early adolescence. Development and Psychopathology. 2010; 22:635–653. http://dx.doi.org/10.1017/S0954579410000337. [PubMed: 20576184]

Myers DE, Milne AM, Baker K, Ginsburg K. Student discipline and high school performance. Sociology of Education. 1987; 60:18–33.

Nelson J, Benner G, Lane K, Smith B. Academic achievement of K-12 students with emotional and behavioral disorders. Exceptional Children. 2004; 71:59–73.

Nelson JR, Benner GJ, Neill S, Stage S. Interrelationships among language skills, externalizing behavior, and academic fluency and their impact on the academic skills of students with ED. Journal of Emotional and Behavioral Disorders. 2006; 14:209–216. http://dx.doi.org/10.1177/10634266060140040401.

No Child Left Behind (NCLB) Act of 2001. Pub. L. No. 107–110, §115, Stat. 1425. 2002.

Panel Study of Income Dynamics. Produced and distributed by the Institute for Social Research, Survey Research Center. University of Michigan; Ann Arbor, MI: 2012. public use dataset

Peterson JL, Zill N. Marital disruption, parent-child relationships, and behavior problems in children. Journal of Marriage and Family. 1986; 48:295–307.

Rabe-Hesketh, S., Skronda, A. Multilevel and longitudinal modeling using Stata. 3rd ed.. Stata Press; College Station, TX: 2012.

Rabiner DL, Murray DW, Schmid L, Malone PS. An exploration of the relationship between ethnicity, attention problems, and academic achievement. School Psychology Review. 2004; 33:498–509.

Ready DD, LoGerfo LF, Burkam DT, Lee VE. Explaining girls' advantage in kindergarten literacy learning: Do classroom behaviors make a difference. The Elementary School Journal. 2005; 106:21–38. http://dx.doi.org/10.1086/496905.

Reid R, Gonzalez J, Nordness P, Trout A, Epstein M. A meta-analysis of the academic status of students with emotional/behavioral disturbance. Journal of Special Education. 2004; 38(3):130–143. http://dx.doi.org/10.1177/00224669040380030101.

Riglin L, Petrides K, Frederickson N, Rice F. The relationship between emotional problems and subsequent school attainment: A meta-analysis. Journal of Adolescence. 2014; 37(4):335–346. [PubMed: 24793380]

Rimm-Kaufman SE, Curby TW, Grimm KJ, Nathanson L, Brock LL. The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. Developmental Psychology. 2009; 45(4):958–972. [PubMed: 19586173]

Rudasill KM, Gallagher KC, White JM. Temperamental attention and activity, classroom emotional support, and academic achievement in third grade. Journal of School Psychology. 2010; 48(2): 113–134. http://dx.doi.org/10.1016/j.jsp.2009.11.002. [PubMed: 20159222]

Rudasill KM, Rimm-Kaufman SE. Teacher–child relationship quality: The roles of child temperament and teacher–child interactions. Early Childhood Research Quarterly. 2009; 24(2):107–120.

Sawka KD, McCurdy BL, Mannella MC. Strengthening emotional support services an empirically based model for training teachers of students with behavior disorders. Journal of Emotional and Behavioral Disorders. 2002; 10(4):223–232.

Shumow L, Vandell DL, Posner JK. Harsh, firm, and permissive parenting in low-income families relations to children's academic achievement and behavioral adjustment. Journal of Family Issues. 1998; 19(5):483–507. http://dx.doi.org/10.1177/019251398019005001.

Spectrum K12. Response to intervention (RTI) adoption survey 2009. Spectrum K12 School Solutions; Towson, MD: 2009.

StataCorp. Stata statistical software: Release 12. StataCorp LP; College Station, TX: 2011.

Stanger C, Lewis M. Agreement among parents, teachers, and children on internalizing and externalizing behavior problems. Journal of Clinical Child Psychology. 1993; 22(1):107–116.

Stormshak EA, Bierman KL, McMahon RJ, Lengua LJ. Parenting practices and child disruptive behavior problems in early elementary school. Journal of Clinical Child Psychology. 2000; 29(1): 17–29. http://dx.doi.org/10.1207/S15374424jccp2901_3. [PubMed: 10693029]

Sugai G, Horner RH. Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered system approaches. Exceptionality. 2009; 17:223–237. http://dx.doi.org/10.1080/09362830903235375.

Valiente C, Lemery-Chalfant K, Castro KS. Children's effortful control and academic competence: Mediation through school liking. Merrill-Palmer Quarterly. 2007; 53(1):1–25.

Vanneman, A., Hamilton, L., Baldwin Anderson, J., Rahman, T. Achievement gaps: How Black and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress, (NCES 2009-455). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education; Washington, DC: 2009.

Vaughn, MG., DeLisi, M., Matto, HC. Human behavior: A cell to society approach. John Wiley & Sons; Hoboken, NJ: 2014.

Verhulst FC, Akkerhuis GW. Agreement between parents' and teachers' ratings of behavioral/emotional problems of children aged 4–12. Journal of Child Psychology and Psychiatry. 1989; 30(1):123–136. [PubMed: 2925818]

Verhulst FC, Koot HM, Van der Ende J. Differential predictive value of parents' and teachers' reports of children's problem behaviors: A longitudinal study. Journal of Abnormal Child Psychology. 1994; 22(5):531–546. [PubMed: 7822627]

Vitaro F, Brendgen M, Larose S, Tremblay RE. Kindergarten disruptive behaviors, protective factors, and educational achievement by early adolescence. Journal of Educational Psychology. 2005; 97:617–629. http://dx.doi.org/10.1037/0022-0663.97.4.617.

Wallace JM, Goodkind S, Wallace CM, Bachman JG. Racial, ethnic, and gender differences in school discipline among U.S. high school students: 1991–2005. Negro Education Review. 2008; 59:47–62.

Weaver-Hightower M. The "boy turn" in research on gender and education. Review of Educational Research. 2003; 73:471–498. http://dx.doi.org/10.3102/00346543073004471.

Wentzel KR. Does being good make the grade? Relations between academic and social competence in early adolescence. Journal of Educational Psychology. 1993; 85:357–364.

Woodcock, RW., Johnson, MB. Woodcock-Johnson psycho-educational battery-revised. DLM Teaching Resources; Allen, TX: 1989.

Wright JP, Morgan MA, Coyne MA, Beaver KM, Barnes JC. Prior problem behavior accounts for the racial gap in school suspensions. Journal of Criminal Justice. 2014; 42(3):257–266.

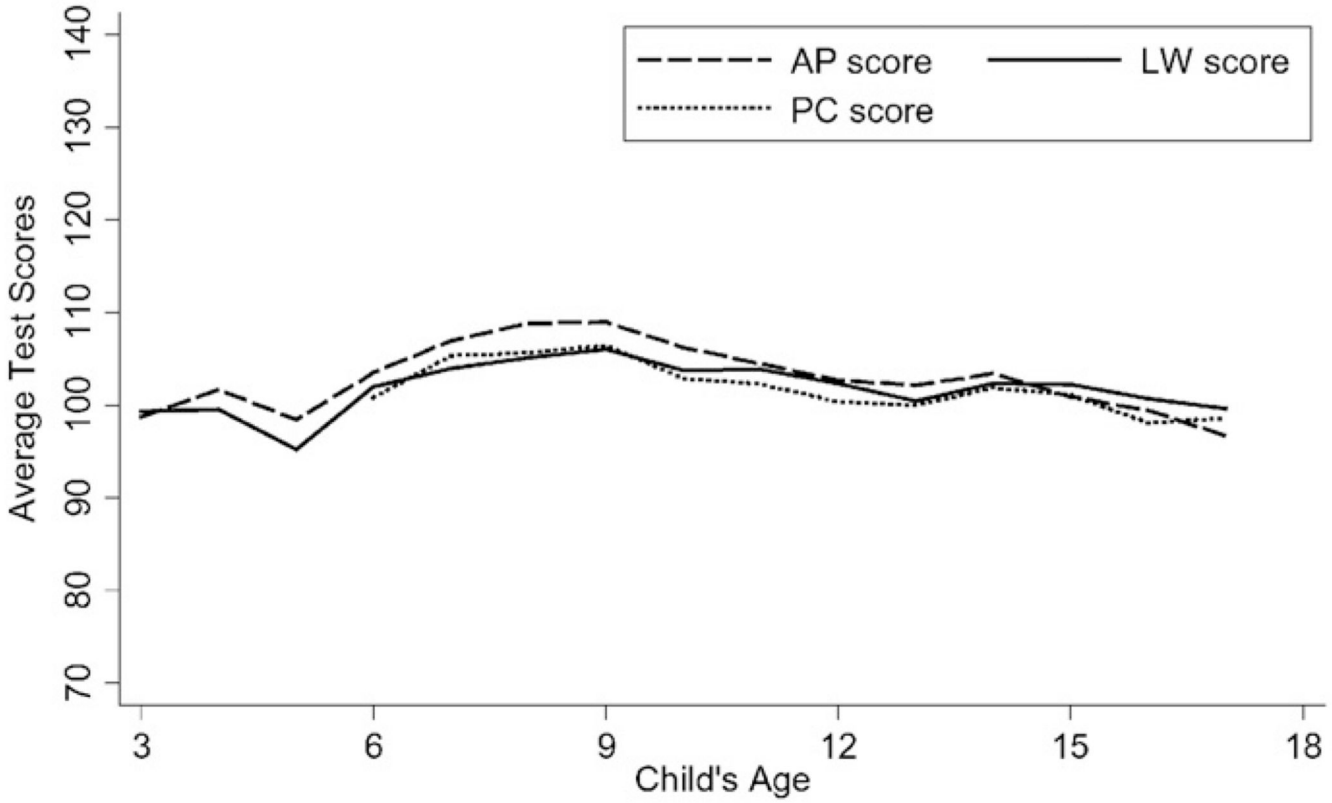

Fig. 1.
Average test scores by child's age.

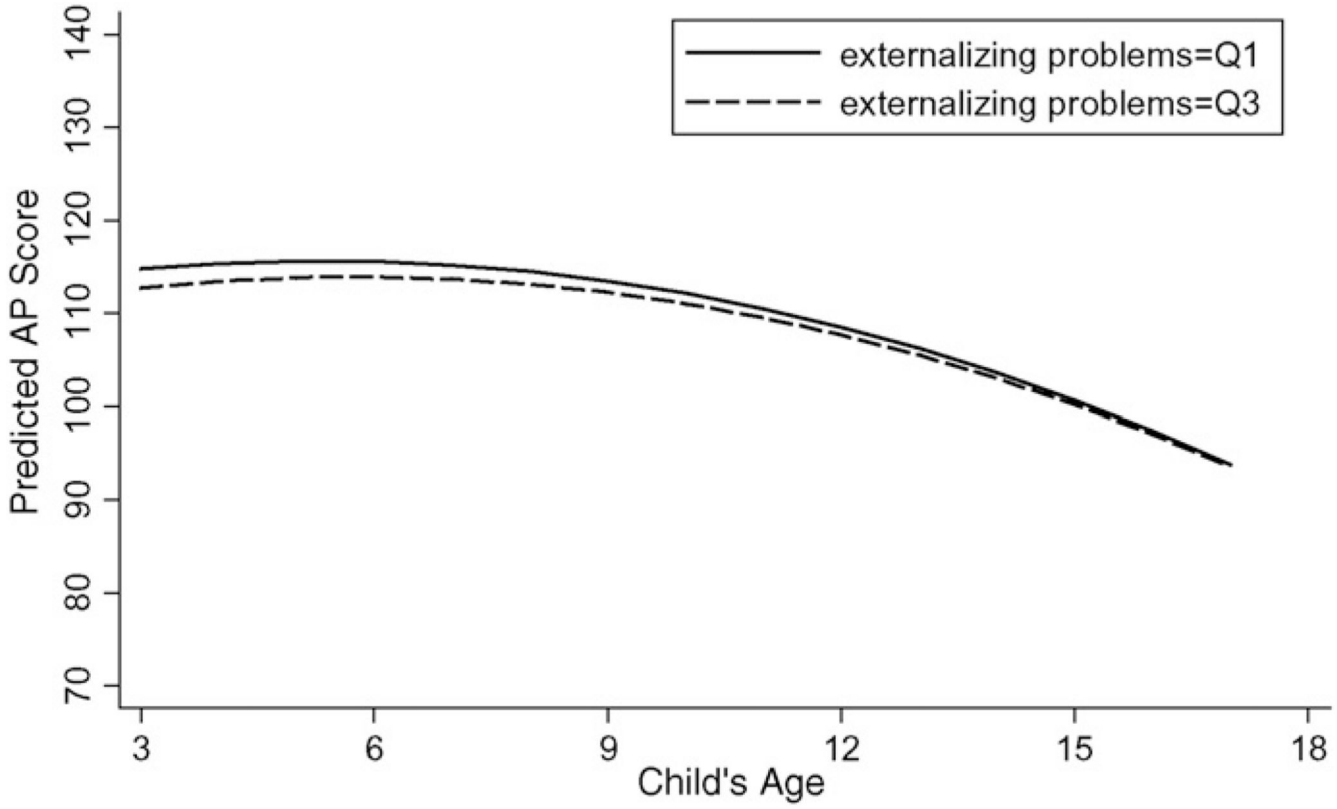

Fig. 2.

Predicted AP scores by child's age and externalizing problems. This figure presents the predicted AP scores over children's age for two typical children in the sample with the externalizing subscale at the third quartile point (= 26) and at the first quartile point (= 19) at baseline. Using the median value of categorical control variables and the mean value of continuous control variables, a typical case is defined as a White male child who had more than 37 weeks of gestational age, had a normal birth weight, did not have neonatal intensive care at birth and did not have physical/mental limitations; whose mother was 36 years old, employed, and had 13 years of schooling; whose mother reported a 4.5 parental warmth score, a 10.6 cognitive simulation score, and a 10.3 emotional support score; whose household had 4.3 members (including 2.4 children) with income equal to $57,000; and whose household did not receive any public assistance.

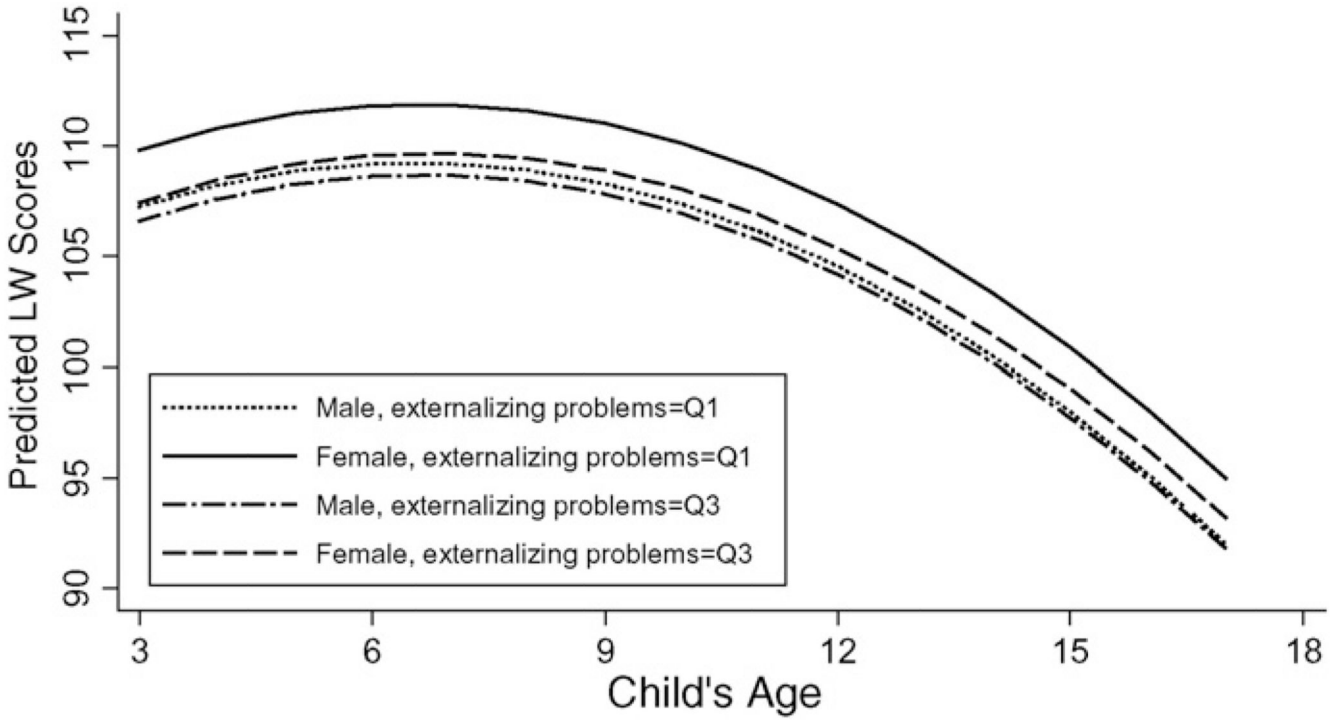

Fig. 3.

Predicted LW scores by child's age, gender, and externalizing problems. Note: This figure presents the predicted LW scores over children's age for four typical children with different genders and with the externalizing subscale at the third quartile point (= 26) and the first quartile point (= 19) at baseline. Using the median value of categorical control variables and the mean value of continuous control variables, a typical case is defined as a White child who had more than 37 weeks of gestational age, had a normal birth weight, did not have neonatal intensive care at birth and did not have physical/mental limitations; whose mother was 36 years old, employed, and had 13 years of schooling; whose mother reported a 4.5 parental warmth score, a 10.6 cognitive simulation score, and a 10.3 emotional support score; whose household had 4.3 members (including 2.4 children) with income equal to $57,000; and whose household did not receive any public assistance.

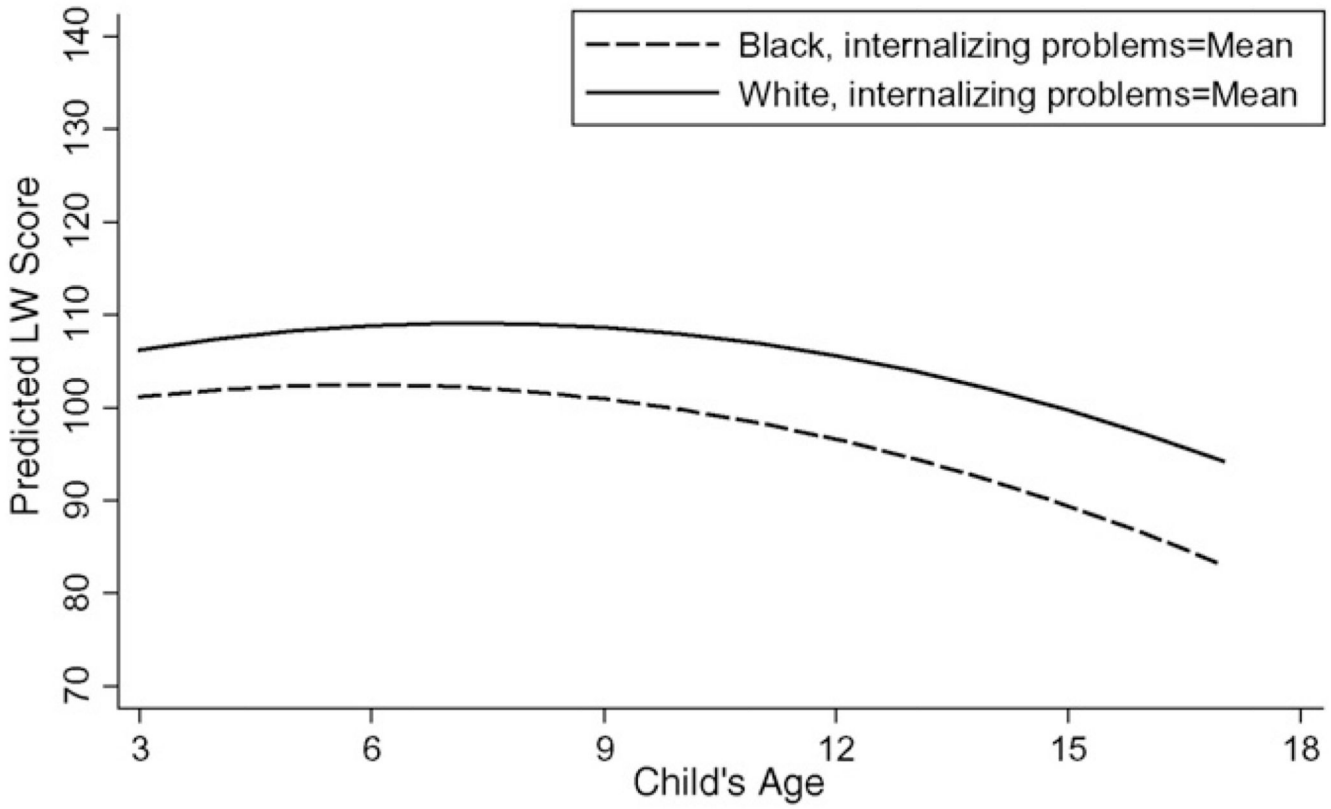

Fig. 4.

Predicted LW scores by child's age, race, and internalizing problems. This figure presents the predicted LW scores over children's age for two typical children with the mean internalizing subscale (= 15) at baseline. Using the median value of categorical control variables and the mean value of continuous control variables, a typical case is defined as a male child who had more than 37 weeks of gestational age, had a normal birth weight, did not have neonatal intensive care at birth and did not have physical/mental limitations; whose mother was 36 years old, employed, and had 13 years of schooling; whose mother reported a 4.5 parental warmth score, a 10.6 cognitive simulation score, and a 10.3 emotional support score; whose household had 4.3 members (including 2.4 children) with income equal to $57,000; and whose household did not receive any public assistance.

Table 1

Characteristics of the analytic sample: PSID-CDS (1997–2007)[a], (*N = 2028*).

Variables[b]	Mean (SD) or %
Three-wave average of test scores	
WJ-R[c] AP Score	107.3 (16.6)
WJ-R[c] LW Score	105.5 (18.4)
WJ-R[c] PC Score	105.2 (17.7)
Major independent variables	
Age in years	8.4 (3.4)
Age range	3–12
Gender (male)	52.7
Race (black)	20.7
Behavior problems	
Externalizing problems	23.1 (5.7)
Internalizing problems	16.0 (3.6)
Covariates	
Children's characteristics	
Preterm birth (yes)	10.4
Low birth weight (yes)	3.1
Neonatal care at birth (yes)	11.7
Learning disability (yes)	10.2
Mothers' characteristics	
Age in years	36.0 (7.7)
Education in years	13.1 (2.7)
Employment status (yes)	65.4
Parental warmth	4.5 (0.5)
Emotional support to children	10.3 (2.0)
Cognitive simulation to children	10.6 (2.0)
Household level	
Household size	4.3 (1.1)
Number of children	2.4 (1.0)
Food stamp participation (yes)	16.4
AFDC[d] participation (yes)	8.0
Household income ($)	57,164.8 (59,227.2)

[a] Panel Study of Income Dynamics – Child Development Supplement.

[b] All variables were measured in the 1997 PSID-CDS except for children's test scores.

[c] Woodcock-Johnson revised.

[d] Aid to families with dependent children.

Table 2

Means of behavior problems and academic achievement by gender and race (N = 2028).

Variables	Gender		Race	
	Female	Male	Black	White
Academic achievement				
WJ-R[a] AP score	105.6	108.9[b]	95.9	110.2[c]
WJ-R LW score	106.9	104.2[b]	94.3	108.3[c]
WJ-R PC score	106.4	104.0[b]	94.2	108.0[c]
Behavior problems				
Externalizing problems	22.5	23.6[b]	23.6	22.9[d]
Internalizing problems	15.9	16.0	16.1	15.9

[a]Woodcock-Johnson revised.

[b]Gender difference significant at the 0.001 level.

[c]Racial difference significant at the 0.001 level.

[d]Racial difference significant at the 0.05 level.

Table 3

Behavior problems and long-term performance of academic achievement.

Variables	Externalizing behaviors			Internalizing behaviors		
	AP score	LW score	PC score	AP score	LW score	PC score
	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)
Intercept (r_{00})	84.03*** (72.52, 95.54)	81.59*** (68.09, 95.09)	90.52*** (78.58,102.47)	82.90*** (71.65,94.15)	76.63*** (63.40, 89.85)	89.49*** (77.78,101.20)
Age (r_{10})	1.36** (0.43, 2.28)	1.99*** (0.99, 2.98)	0.72 (−0.71,2.15)	1.40** (0.60,2.20)	2.10*** (1.23,2.98)	0.44 (−0.90,1.78)
Age squared (β_2)	−0.16*** (−0.21, −0.11)	−0.16*** (−0.21, −0.11)	−0.13* (−0.24, −0.02)	−0.16*** (−0.21, −0.11)	−0.16*** (−0.21, −0.11)	−0.13* (−0.24, −0.02)
Gender (male)	3.02*** (1.92,4.10)	−1.70** (−3.00, −0.40)	−1.36* (−2.54, −0.19)	2.87*** (1.78, 3.96)	−1.84** (−3.14, −0.53)	−1.53** (−2.71, −0.36)
Race (Black)	−10.88*** (−12.43, −9.34)	−7.56*** (−9.43, −5.69)	−7.35*** (−8.99, −5.71)	−10.94*** (−12.50, −9.39)	−7.50*** (−9.38, −5.61)	−7.37*** (−9.02, −5.70)
Behavior problems	−0.35*** (−0.53, −0.17)	−0.22* (−0.40, −0.03)	−0.22* (−0.39, −0.05)	−0.50*** (−0.80, −0.20)	−0.12 (−0.44,0.20)	−0.35** (−0.64, −0.06)
Behavior problems * age	0.02* (0.00, 04)	0.00 (−0.01,0.02)	−0.00 (−0.02, 0.02)	0.03 (−0.00, 05)	−0.00 (−0.03, 0.03)	0.02 (−0.02, 0.06)
Number of children	2028	2028	1960	2028	2028	1960
Number of observations	4119	4119	3.591	4119	4119	3.591

a. All three models are adjusted for the following covariates: (1) children's race, gender, preterm birth, low birth weight, neonatal intensive care, and physical/mental limitation; (2) mother's age, education, employment status, parental warmth, emotional support, and cognitive stimulation; and (3) household's size, number of children, food stamp participation, Aid to Families with Dependent Children or Temporary Assistance for Needy Families participation, homeownership, household income, and state fixed effects. b. The analysis on the PC score has a smaller sample size since only children aged 6 or older can take this test

* $p < 0.05$.

** $p<0.01$.

*** $p< 0.001$.

Table 4

Behavior problems and long-term performance of academic achievement by gender.

Variables	Externalizing behaviors			Internalizing behaviors		
	AP score	LW score	PC score	AP score	LW score	PC score
	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)
Behavior problems * male	0.02 (−0.18, 0.23)	0.25 * (0.02, 0.48)	0.20 (−0.02, 0.42)	−0.06 (−0.38, 0.27)	0.31 (−0.09,0.70)	0.15 (−0.19, 0.50)
Behavior problems * age * male	0.00 (−0.01,0.01)	−0.00 (−0.01,0.01)	−0.01 (−0.02,0.01)	−0.00 (−0.01,0.01)	−0.00 (−0.02, 0.01)	−0.00 (−0.02, 0.01)
Number of children	2028	2028	1960	2028	2028	1960
Number of observations	4119	4119	3.591	4119	4119	3.591

a. All three models are adjusted for the following covariates: (1) children's race, gender, preterm birth, low birth weight, neonatal intensive care, and physical/mental limitation; (2) mother's age, education, employment status, parental warmth, emotional support, and cognitive stimulation; and (3) household's size, number of children, food stamp participation, Aid to Families with Dependent Children or Temporary Assistance for Needy Families participation, homeownership, household income, and state fixed effects. b. The analysis on the PC score has a smaller sample size since only children aged 6 or older can take this test.

* $p < 0.05$.

Table 5

Behavior problems and long-term performance of academic achievement by race.

Variables	Externalizing behaviors			Internalizing behaviors		
	AP score	LW score	PC score	AP score	LW score	PC score
	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)	b (95% CI)
Behavior problems * Black	0.02 (−0.19, 0.23)	−0.05 (−0.29, 0.19)	−0.10 (−0.32, 0.12)	−0.10 (−0.43,0.24)	−0.24 (−0.65, 0.17)	−0.10 (−0.32, 0.12)
Behavior problems * age * Black	−0.01 (−0.01,0.00)	−0.02*** (−0.03, −0.01)	−0.02** (−0.03, −0.01)	−0.00 (−0.02,0.00)	−0.03*** (−0.04, −0.01)	−0.02** (−0.03, −0.01
Number of children	2028	2028	1960	2028	2028	1960
Number of observations	4119	4119	3.591	4119	4119	3.591

a. All three models are adjusted for the following covariates: (1) children's race, gender, preterm birth, low birth weight, neonatal intensive care, and physical/mental limitation; (2) mother's age, education, employment status, parental warmth, emotional support, and cognitive stimulation; and (3) household's size, number of children, food stamp participation, Aid to Families with Dependent Children or Temporary Assistance for Needy Families participation, homeownership, household income, and state fixed effects. b. The analysis on the PC score has a smaller sample size since only children aged 6 or older can take this test.

** $p < 0.01$.

*** $p < 0.001$.