The Relations between Academic Achievement and Externalizing Behavior: Separating Fact from Fiction

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA
BY

Tara Kulkarni

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Advisor: Amanda L. Sullivan, PhD, LP

Expected Date of Graduation: June 2021
Acknowledgements

I want to thank my family, my advisor Dr. Amanda Sullivan, and my committee members Drs. Annie Hansen-Burke, Amy Hewitt, and Nidhi Kohli for providing me with the support, encouragement, and constructive feedback needed to bring this document to life. Thank you to all of my friends for keeping my spirits up these last few months. Most of all, thank you to everyone who believed in me. I moved literal continents to follow my passion, and it has all been worth it.
Abstract

Some of the worst long-term outcomes of children are associated with the presence of both externalizing behavior and low academic achievement. Additionally, trajectories of externalizing problem behavior have shown that when children enter kindergarten with problem behavior, it tends to persist and is often associated with low academic achievement. However, though there remains a popular belief amongst educators that academic achievement and externalizing behavior have a strong predictive and even causal relationship, evidence is mixed. Given the implications for both resource allocation and intervention design if causal associations were supported, this dissertation sought to examine the relationship between the two domains, by (a) systematically reviewing literature in an effort to reveal potential causal relations, if any, and (b) conduct an empirical study using nationally representative data (N=7,330) and latent class growth analysis to reveal relations of early academic achievement with externalizing behavior trajectories based on the findings of the review. Results from both studies indicated that there is no concrete evidence for even predictive relations between achievement and externalizing behavior. Instead, the low achievement often observed in children with high externalizing behavior likely has other underlying causes. Specifically, results suggested that malleable variables like inattention and school readiness behaviors are better predictors of both achievement and teacher reported externalizing behavior.

Lastly, this dissertation also revealed that socio-demographic factors like sex and race have strong associations with teacher reported externalizing behavior. Implications for school systems as well as student level interventions are discussed.
Table of Contents

Acknowledgements ........................................................................................................................................ iii

Abstract .................................................................................................................................................... iv

Table of Contents ........................................................................................................................................ iv

Chapter 1: Introduction ............................................................................................................................... 1

Evidence of Association ............................................................................................................................... 4

Chapter 2: Externalizing Behavior Problems and Low Academic Achievement:

   Does a Causal Relationship Exist? ........................................................................................................... 8

   Evidence and Theories of Association ...................................................................................................... 9

   Current Gaps in Literature .................................................................................................................... 12

   Establishing Causal Associations ......................................................................................................... 14

   Method .................................................................................................................................................... 16

   Results .................................................................................................................................................... 8

      Sample Characteristics .......................................................................................................................... 21

      Study Design Characteristics .............................................................................................................. 22

   Does Evidence Support a Causal Relationship Between Externalizing Behavior Problems and Low Academic Achievement? .............................................................. 23

       What Are the Mechanisms Moderating or Mediating the Relationship? ........................................ 23

      Summary and Quality Analysis .......................................................................................................... 25

   Discussion .............................................................................................................................................. 26

      Heterogeneity in Research Design ....................................................................................................... 29

      Implications for Practice ................................................................................................................... 31

      Limitations ......................................................................................................................................... 32
Table of Contents

Chapter 3: Early Academic Achievement and Relations to Trajectories of

Externalizing Problem Behavior in Elementary School ........................................33

Method ......................................................................................................................39

Results .....................................................................................................................57

Model Selection .....................................................................................................57

Three Class Model .................................................................................................63

Research Question 1 .............................................................................................64

Research Question 2 .............................................................................................65

Discussion ...............................................................................................................68

Early Academic Achievement and Externalizing Behavior .............................69

Socio-Demographic Characteristics ......................................................................73

Implications for Practice .......................................................................................76

Limitations ...............................................................................................................79

Chapter 4 Discussion .............................................................................................81

Implications for Practice .......................................................................................83

Future Directions ..................................................................................................86

Conclusion .............................................................................................................87

References ..............................................................................................................89

Appendix A ............................................................................................................152

Quality Indicator Checklist (Tooth et al., 2005) ..................................................152

Appendix B ............................................................................................................153

Search Terms ..........................................................................................................153

Appendix C ............................................................................................................155
Table of Contents

Train and Test Data Set ........................................................................................................155

List of Tables and Figures

Table 1.1 Descriptives of Studies Included in Systematic Review ...............................121
Table 1.2 Studies that Examined Bidirectional Models ..............................................122
Table 1.3 Studies that Examined Unidirectional Models ........................................ 126
Table 1.4 Study Quality Scores and Results by Statistical Design .............................131
Table 1.5 Quality Indicator Scores ..............................................................................133
Table 2.1 Descriptives of Variables in Analysis ..........................................................135
Table 2.2 Proportion of Zeros in Each Wave of Data ...............................................137
Table 2.3 Correlation Matrix .......................................................................................138
Table 2.4 Fit Statistics for One, Two, and Three Class Models ................................140
Table 2.5 Mean Parameters Estimates of Models .......................................................141
Table 2.6 LCGA Regression Results ...........................................................................142
Table 2.7 Between Class Mean Comparison ................................................................143

Figure 1.1. Visual representation of potential causal pathways between
externalizing behavior problems and academic achievement ............................144

Figure 1.2. Mediation model between academic achievement and
externalizing behavior problems ..........................................................................145

Figure 1.3. PRISMA Chart ............................................................................................145

Figure 2.1. Distribution frequencies of outcome variable at all waves ......................148

Figure 2.2 Mean teacher rated scores of externalizing behaviors from
kindergarten to fifth grade .........................................................................................149
Table of Contents


ded by the author's name.

Figure 2.3 Mean teacher ratings of externalizing behavior through fifth grade

for two and three class solutions respectively........................................151
Chapter 1: Introduction

Some of the worst long-term outcomes of children are associated with the presence of both externalizing behavior and low academic achievement (e.g. Bushman et al., 2018; Morgan, Farkas, & Wu, 2012). Additionally, trajectories of externalizing problem behavior have shown that once children enter kindergarten with problem behavior, it tends to persist and is often intractable (Olson, Choe, & Sameroff, 2017). Although there are several factors associated with the development of problem behavior and academic difficulties, studies have shown that their association is such that success in one domain might protect against failure in the other (Masten et al., 2005). Thus, intervention in one domain could prevent development of difficulties in the other, mitigating risks of negative outcomes associated with dual failures (Bradshaw, Zmuda, Kellam & Ialongo, 2009). However, even though many studies have reported associations between the two, evidence for both the causal and predictive association has been inconsistent (Algozzine, Wang & Violette, 2011; Hinshaw, 1992).

Identifying whether the association between externalizing and achievement domains is reliable and valid could inform prevention practices to reduce problem behavior and also boost academic achievement (Duncan et al., 2007). Additionally, given that multiple studies have shown that groups of students exhibit different trajectories of externalizing behavior in school (e.g. Nagin & Tremblay, 1999; Olson et al., 2017), identifying whether the association between achievement and behavior is strongest for a particular subpopulation of children could improve targeted prevention practices (Reinke, Herman, Petras, & Ialongo, 2008). Thus, the purpose of this dissertation is two part: (1) to examine if current best evidence supports a causal relationship between academic
achievement and externalizing behavior and describe potential mediators and moderators of the relationship; and (2) to conduct a secondary data analysis that examines whether a student’s early academic achievement is associated with membership to differential behavior trajectories in elementary school, and to what extent child or school level characteristics contribute to this relationship.

**Development of Comorbid Academic and Behavioral Difficulties**

Externalizing behavior problems are characterized by behavior that is aggressive, hyperactive, or disruptive (Achenbach, 1966; 1999). Children with disruptive or externalizing behavior problems have some of the worst academic outcomes including higher rates of drop out (Piquero, Jennings, & Barnes, 2012), lower rates of academic engagement, and poor student-teacher relationships (Sabol & Pianta, 2012). Dynamic systems theory (Sameroff, 2000) suggests that changes in one area of functioning can trigger a sequence of consequences that ultimately have large developmental effects (Sameroff, 2000), or a developmental cascade (Cicchetti & Masten, 2010). For example, if a child enters school with significant disruptive behavior, the resulting poor interactions with peers and teachers could reinforce problematic behavior and ultimately lead to less opportunities to experience both social and academic success (Patterson, DeBaryshe & Ramsey, 2017), leading to the observed comorbidity between externalizing behavior and low academic achievement.

This comorbidity could also develop by repeated failure in academic tasks (Patterson et al., 2017). To avoid expected failure and resultant feelings of frustration, children may learn that engaging in maladaptive behaviors, commonly disruptive, allows them to escape the uncomfortable or undesired academic task (McIntosh, Horner, Chard,
Boland & Good, 2006). Additionally, children with externalizing behavior symptoms are likely to face a higher number of school disciplinary actions including suspension, expulsion, and subsequently reduced instruction time, access to general education and eventually low academic engagement and achievement (McIntosh et al., 2006). This comorbid failure, in behavior and academic achievement, is believed to have bidirectional effects as children complete their school career and can lead to higher probabilities of dropout and even adult psychopathology (Masten et al., 2005). Thus, comorbidity could develop by either trigger; academic as well as behavioral difficulties.

Lastly, though most research has focused on how psychopathology develops within the family unit, neither child behavior nor academic achievement exists in the vacuum of parent-child interactions (Bronfrenbrennor, 1994). Though research has shown that parents who experience high amounts of stress are the least likely to implement positive behavioral strategies consistently (Aizer, Stroud & Buka, 2016), these parents are also overwhelmingly likely to experience poverty and have the least access to resources, contributing directly to their levels of stress and mental health (Neppl, Senia & Donellan, 2016). In the schools, decades of research have indicated a child’s identification for special education services, especially in the category of emotional and behavior disorders, which serves mostly students with comorbid academic and behavioral difficulties (Wagner et al., 2004) can be associated with their race (Skiba et al., 2016). Or rather, identification for special education can be viewed within the framework of complex interactions of systemic inequities and racism faced by persons of color (Annamma, Morrison, & Jackson, 2014). Additionally, public schools that have the fewest resources often have the worst outcomes in terms of behavior and academic
achievement (Sullivan, Klingbeil & Van Norman, 2013), with a higher number of children scoring in the bottom percentile in state test scores, being identified for special education, and entering into the juvenile justice system (Duncan & Murnane, 2011; Skiba & Rausch, 2004). Thus, the development of difficulties in either domain is a result of a complex set of variables that include child, parent, school, and environment conditions (Bierman et al., 2013; Dodge & Haskins, 2015).

**Evidence of Association**

Multiple studies have examined the association between achievement and externalizing problem behavior, however, results have not been consistent (e.g. Hinshaw 1992; Grimm, Steele, Mashburn, Burchinal & Pianta, 2010). Though cross-sectional studies have documented the presence of comorbid difficulties in children (e.g., Volpe, 2006), drawing conclusions from these designs is challenging due to the inability of parsing whether behavior or academic difficulties precede the other (Kraemer, 2001). Several studies with longitudinal designs, which allow for predictive inferences, have reported that early behavioral difficulties in elementary school are negatively associated with academic achievement in middle and high school (e.g. Masten et al., 2005; Dishion, Patterson, Stoolmiller & Skinner, 1991; vanLier & Koot, 2010), whereas a few have reported similar negative associations between academic difficulties and later externalizing behavior (Burt & Roisman, 2010; Masten et al., 2005). This predictive relationship has also been examined in very young children, where maternal report of early externalizing or aggressive behavior in two to three years olds’ was associated with poorer academic achievement in first and second grade (e.g. Bub, McCartney & Willet, 2007; Grey et al., 2014).
Other studies have suggested that gender and race could have a moderating effect on the association between the two domains (e.g. Kremer, Flower, Huang & Vaughn, 2017). There have also been several studies that have reported little to no relationship between the two domains across developmental stages (e.g. Algozzine et al., 2011; Burt & Roisman, 2010; Duncan et al., 2007; Massetti et al., 2008). Additionally, several have reported moderate associations of socioeconomic status (Breslau et al., 2009; Morgan et al., 2018) and cognitive ability (Burt & Roisman, 2010; Miller et al., 2014) on the outcomes of both achievement and behavior. Differential conclusions of articles examining the same association are therefore theorized to be due, in part, to heterogenous study designs, measurement error, as well as inadequate statistical controls (e.g. Algozzine et al., 2011, Grimm et al., 2010, Hinshaw, 1992).

Even though randomized experimental studies, that allow causal inferences, examining the relationship between the externalizing behavior problems and academic achievement are challenging due to the inability to manipulate the primary variables (Schneider et al., 2007), close approximations to randomized trials can inform prevention and intervention efforts (Kraemer, 2001). Several school- and class-wide experimental studies targeting externalizing and disruptive behavior have examined academic achievement outcomes associated with the implementation of the interventions. At the school-level, a recent meta-analysis of universal social and emotional learning programs designed to reduce problem behavior and improve prosocial behavior, found that the relationship between the intervention and academic achievement was strongly correlated with study quality, and that high-quality randomized studies did not show meaningful effects (Corcoran, Cheung, Kim and Xie, 2018). Similarly, systems-level efforts to
reduce reports of problem behavior behavior, like School Wide Positive Behavior Intervention and Supports [SWPBIS], have shown consistent positive effects on reducing problem behavior, but mixed results on improved academic achievement (Noltemeyer, Palmer, James & Wiechman, 2018), conclusions which did not differ based on the implementation level of SWPBIS (James, Noltemeyer, Ritchie & Palmer, 2019).

Even though evidence regarding the association is mixed, research continues to demonstrate that the comorbidity between problems is common (e.g., Tannock et al., 2018). The mechanisms through which this association develops are complex and implicate several variables (Hinshaw, 1992; Masten et al., 2005), however given the negative outcomes associated with problems in both domains (Reinke, Herman, Petras & Ialongo, 2008), the ability to validate a causal association between the two could improve targeted intervention practices, especially in schools (Bradshaw et al., 2009; McIntosh et al., 2006). This is especially important for early intervention, as intervening on a causal mechanism when children are younger could have cascading impacts on longitudinal outcomes that are related to the deficits (Cicchetti & Masten, 2010). Moreover, as research has demonstrated that there are multiple problem behavior trajectories (e.g. Nagin & Tremblay, 1999), using a person-centered approach rather than a variable-centered approach could help parse whether the association between achievement and behavior difficulties is stronger for certain groups of children (Reinke et al., 2008; Walrath et al., 2004). Thus, this paper will attempt to answer the following questions in two studies:
1. (a) Does available empirical literature support a causal association between academic achievement and externalizing behavior problems, and, if so, in what direction?
   (b) What are the mechanisms moderating or mediating this relationship?
2. Is a student’s early academic achievement associated with membership to differential developmental behavioral trajectories, and what are the child, family and school characteristics that are associated with this relationship?
Chapter 2: Externalizing Behavior Problems and Low Academic Achievement: Does a Causal Relationship Exist?

As early as kindergarten, externalizing behavior problems and low academic achievement are associated with negative outcomes, both proximal and distal (e.g., Bushman et al., 2018; Morgan, Farkas, & Wu, 2009). Additionally, both domains are highly correlated (e.g., Murray & Farrington, 2010) such that, together, difficulties severely undermine a child’s academic and mental health outcomes (Nelson, Benner, Lane, & Smith, 2004). Children who present with concomitant academic failure and externalizing behavior problems often have some of the worst life outcomes as adults, including incarceration (Moffit, 2003). Although a large body of studies indicate a predictive relation between the two (Algozzine, Wang, & Violette, 2011), establishing whether the evidence supports a causal association is essential to inform effective intervention for prevention of persistent academic failure and psychopathology (Masten & Cicchetti, 2010) and streamlining of efficient resource allocation in schools (McIntosh, Horner, Chard, Boland & Good, 2006). The only other review examining causal mechanisms between externalizing behavior and low academic achievement reported a modest association between the two domains, but the generalizability of results was hindered by the lack of studies with appropriate methodology (Hinshaw, 1992).

The purpose of this study, therefore, was to update Hinshaw’s (1992) review by systematically reviewing the empirical research on the association between low academic achievement and externalizing behavior problems, examine evidence for causal mechanisms, if any, between the two domains, and report on any mediators or moderators of the relation. The term low or poor achievement is used in this review instead of
underachievement in an effort to include the relevant population of children who have academic deficits and not academic discrepancies\(^1\). Similarly, externalizing behavior is defined as first proposed by Achenbach (1966) who described three subtypes within the broader externalizing domain: hyperactive, aggressive, and delinquent.

**Evidence and Theories of Association**

Early academic skills are some of the best predictors of not only future academic achievement but also positive peer relationships, student-teacher relationships, and school adjustment (e.g., Hall, Simon, Lee & Mercy, 2012; Sabol & Pianta, 2012; Spira & Fischel, 2005). Conversely, poor achievement in young children is associated with increased adolescent behavioral difficulties (e.g., Darney, Reinke, Herman, Stormont, & Ialongo, 2012), juvenile delinquency (e.g., Piquero, Jennings, & Barnes, 2012), and adult psychopathology (e.g., Patterson, DeBaryshe, & Ramsey, 2017). Not only do poor achievement and externalizing behavior problems often co-occur (e.g., Bradshaw, Buckley & Ialongo, 2008; Pardini & Fite, 2010), but each may act as a cumulative risk factor for the other, consequently worsening both short term and long-term outcomes (Murray & Farrington, 2010). Understanding the potential cause-effect relations between these two domains thus becomes critical to designing effective prevention efforts (Masten, Burt & Coatsworth, 2015).

Hinshaw (1992) made the first attempt to review the causal mechanisms between externalizing behavior problems and low academic achievement. He surmised that there were four pathways through which the relation could exist. The first two pathways were

---

\(^1\) Theoretically, underachievement is defined as achievement that is discrepant from a student’s potential as measured by intelligence tests (Zeigler, Zeigler & Stoeger, 2012). Thus, often encompasses children whose scores on achievement tests are average or above average but are discrepant from their expected performance.
unidirectional (i.e., either low achievement preceded and caused externalizing behavior problems or externalizing behavior problems preceded and caused low achievement). The third pathway was bidirectional wherein both domains exerted influence on each other in parallel; and lastly, the fourth pathway hypothesized that an antecedent variable, like language, could precede and cause both domains (Figure 1). Most pathway theories of the development of problem behavior are situated within the dynamic systems theory, which suggests that changes in one area of functioning can trigger a sequence of consequences that ultimately have large developmental effects (Sameroff, 2000), or a developmental cascade (Masten & Cicchetti, 2010). A brief overview of theories put forth for each potential pathway is presented below.

**Unidirectional and bidirectional pathways.** Some evidence supports a unidirectional pathway (Figure 1a) in which externalizing behavior problems in early childhood precede, and consequently negatively influence, academic achievement (Capaldi, Jamerson & Patterson, 1997). A dual failure across domains emerges over time as academic failure reduces access to many pathways to positive change (Moffit et al., 2002) and consequently increases the likelihood of negative outcomes (Patterson, Debaryshe & Ramsey, 1990). Conversely, Dishion, Patterson, Stoolmiller, and Skinner (1991) posited that academic difficulties could also precede, and therefore trigger the cascade; i.e., academic failure leads to feelings of frustration and incompetence in children, which exacerbates both externalizing and internalizing symptoms. Thus, the direction of influence could be bidirectional (Figure 1b) and each unidirectional pathway could theoretically exist in parallel; i.e., problems in either domain could cause a worsening performance in the other (Miles & Stipek, 2006). These three theoretical
models were often thought of as too simplistic and thus a fourth model was also theorized (Hinshaw, 1992).

**Antecedent variable pathway.** Hinshaw’s fourth causal hypothesis was that of an antecedent variable: an underlying variable which could explain or partially explain the association between low achievement and problem behavior (Figure 1c). These theoretical antecedents could be child-level; language impairment, for example, is robustly associated with both externalizing behavior problems (e.g., Chow & Wehby, 2018) and deficits in reading (Dickinson, Golinkoff, & Hirsh-Pasek, 2010) and math (Peng et al., 2020). Intellectual ability, commonly referred to as intelligence quotient (IQ), is another potential antecedent, with several scholars theorizing that it could play a key causal role in both low achievement and behavior problems (Fergusson & Horwood, 1995; Hinshaw, 1992; Murray & Farrington, 2010). Antecedents could also be environmental. Low socioeconomic status (SES), for example, has an inconsistent association with problem behavior (e.g., Quan & McGrath, 2014), but its positive association with achievement is more reliable (Lam, 2014; Sirin, 2005). Early parenting practices have also been implicated in the development of both problem behavior and academic achievement, and researchers argue that responsive, warm, and authoritative parenting can predict child success in academic as well as behavioral domains (e.g., Grietens et al., 2004). Controlling for these potential antecedent causes thus becomes critical if examining the relations between externalizing behavior problems and academic achievement (Hinshaw, 1992); however, antecedent variables are only one kind of confound, and other theoretical factors shown to be associated with both achievement and externalizing behavior problems are also important to consider (Metsäpelto et al., 2015).
Other potential mediators. The relation could also be at least partially explained by a potential third variable, specifically, inattention (e.g., Hinshaw, 1992; Fergusson & Horwood, 1995). This theory, first tested by Fergusson and Horwood (1995), suggested that academic deficits do not stem from disruptive or externalizing behavior problems but from inattentive behaviors. For example, some reviews have reported that, when inattention and externalizing behavior problems are examined simultaneously, attention problems predict math and reading achievement consistently, regardless of age, whereas the influence of externalizing problems on achievement is significantly reduced (e.g., Polderman, Boomsma, Bartels, Verhulst, & Huizink, 2010). However, others have reported that externalizing behavior problems remained a robust predictor of academic achievement even after accounting for inattention (Trzesniewski, Moffitt, Caspi, Taylor, & Maughan, 2006). Overall, conclusive evidence remains elusive (Algozzine et al., 2011), suggesting that the association between low achievement and externalizing behavior problems may not be a simple pathway and could have several potential antecedents and mediators (Hinshaw, 1992).

Current Gaps in Literature

Given the negative outcomes associated with both externalizing behavior problems and low achievement, an accurate representation of the relation becomes critical to informing prevention efforts (McIntosh et al., 2006). Meta-analyses and systematic reviews prove vital in providing these accurate representations as they summarize literature in a way that satisfies two criteria necessary to validate causal associations: consistency of findings and strength of association (Weed, 2000). Recently, Savage, Ferguson, and Flores (2017) published a meta-analysis on the effects of
academic achievement on physical aggression, but their review was limited to only those studies that examined violent behavior, and no other dimensions of externalizing behavior like hyperactivity, oppositionality, or non-violent acts of aggression. They also included cross-sectional designs in their study, thus the consequent lack of temporal precedence hindered any potential causal inferences (Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007).

Additionally, Hinshaw’s (1992) own conclusions were overwhelmingly cautionary given the lack of studies with appropriate methodology. He reported that limitations in study designs, including insufficient testing of all possible causal models (Figure 1), unstandardized variables, and inappropriate statistical methodology hindered any reliable causal inferences. Others have stated that the measurement error inherent in quantifying both low academic achievement and externalizing behavior could also influence study outcomes (Algozzine et al., 2011). Thus, even though multiple studies have examined the predictive and causal relations between poor achievement and externalizing behavior problems since Hinshaw’s 1992 review (e.g., Duncan et al., 2007; Masten et al., 2005), to our knowledge, there has yet to be an attempt to systematically review this research with the goal of revealing potential causal associations.

Establishing Causal Associations

Investigating the cause-effect relation between low achievement and problem behavior can have practical applications in education and psychology, especially in schools (Hinshaw, 1992). Schools could improve targeted intervention efforts on the causal mechanism to enhance both academic and behavioral functioning (Bradshaw, Zmuda, Kellam, & Ialongo, 2009; McIntosh et al., 2006), and as a result, broader
outcomes shown to be related to deficits (Masten & Cicchetti, 2010). Thus, intervention in one domain could prevent development of problems in the other, potentially saving schools personnel and financial resources (Bradshaw et al., 2009; van Lier & Koot, 2010). Unfortunately, when examining psychological constructs like behavior and academic achievement, providing evidence for cause and effect becomes challenging (Murray, Farrington & Eisner, 2009).

First, achievement or externalizing behavior could have a multitude of factors correlated with them, making each just one of many correlates that are associated with an outcome (Offord & Kraemer, 2000). To be considered causal, however, apart from establishing temporal precedence, a construct, when manipulated or changed, must also show a corresponding change in the outcome (see Kramer et al., 2001 for a more detailed review) and show change relative to a control condition or alternative cause (Schneider et al., 2007). Randomized experiments are considered the gold standard to infer this cause-effect relation, but in fields like education and social science, it is often neither feasible nor ethical to design true randomized clinical trials (Schneider et al., 2007). Even though scientific manipulation of factors like externalizing behavior problems may not be possible, researchers argue that quasi-experimental and observational studies are able to provide valid evidence, e.g., research designs that approximate conditions that mimic requirements of a randomized clinical trial, such as propensity score analysis or regression discontinuity designs (see Shaddish, 2010 for a more detailed discussion).

A second challenge in establishing causal associations between psychological constructs like externalizing behavior problems and achievement is that they can theoretically be measured at every timepoint, except in very young children, making
establishing temporal precedence a challenge. To address this, Kramer et al. (2001), in their seminal paper on causal risk factors, suggested dichotomizing variables (e.g., low achievement or high externalizing behavior problems) could help define a “threshold” (p.851), which when crossed enables researchers to establish the presence of maladaptive behavior, thus establishing precedence. Without this dichotomization, Kraemer et al. (2001) argued that predictive results have little significance for practical application.

However, the counter argument is that dichotomizing variables can lead to a loss of full information and therefore biased estimates (Cohen, 1994). Thus, both statistical methodology and measurement of variables could impact parameter estimates and causal inference (Kramer et al., 1999; Finney & DiStefano, 2006).

Yet, despite various challenges, the ability to validate causal associations is integral to scholars interested in improving outcomes, and especially useful to those interested in effectiveness of interventions (Glass, Goodman, Hernen & Semet, 2013). This is because designing interventions that manipulate variables that are just correlates and have not been established to have temporal precedence or a causal association have little use (Kraemer et al., 2001). Additionally, considerations of mediators and moderators can be critical to forming valid conclusions about the strength of the association between two variables (Kazdin, 2007).

The purpose of this research article therefore, is to systematically review studies of the relations between externalizing behavior problems and academic achievement to examine any potential causal relations between the two domains, including discussion of potential mediators and moderators. In addition, this study will attempt to address Hinshaw’s (1992) stated limitations by focusing on studies which (a) are able to examine
multiple causal pathways, (b) use standardized measures, and (c) apply research designs that potentially allow for causal inference. Additionally, unlike Hinshaw’s (1992) analysis, which treated IQ and early academic achievement as interchangeable proxies for cognitive ability, this review regarded them as distinct variables given their distinct natures (Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012). Specifically, two research questions will be answered:

1. Does research support a causal association between poor achievement and externalizing behavior problems, and, if so, in what direction?
2. What are the mechanisms moderating or mediating this relation?

**Method**

**Inclusion Criteria**

Inclusion criteria were: (a) articles or dissertations published after Hinshaw’s (1992) review to December 2018; (b) articles were in English (c) externalizing behavior problems or achievement outcomes were measured; (e) a clear temporal ordering of externalizing behavior and achievement variables existed (Schneider et al., 2007); (d) outcomes were for children aged 18 years or younger; (e) standardized, direct achievement measures were reported (Hinshaw, 1992); and (h) design accounted for IQ or language skills. Articles were excluded if they (a) were cross sectional; (b) did not include baseline or control group if experimental; (c) tested interventions that were school level frameworks like positive behavior intervention supports and not direct interventions designed to reduce externalizing behavior; (d) tested programs like social-emotional learning that targeted several domains of behavior, not only externalizing problem behavior (CASEL, n.d); (e) did not include standardized measures of academic
achievement (e.g. teacher or parent report, grade point averages, self-report); (d) were narrative or theoretical (e.g. systematic reviews, opinion pieces); (e) did not control for intellectual or language ability; (f) did not test the relations between behavior and achievement (e.g. if domains were exclusively outcomes or predictors, or only one domain was included in the analysis); or (g) were not English.

Direct standardized measures of achievement were defined as assessments specifically measuring academic achievement (e.g., reading, math, writing), those with standardized questions and norms, either local or national, and were directly administered to the participants. Studies that used standardized national exam scores (e.g., General Certificate of Secondary Education) as assessments were considered but then excluded if the outcome did not measure the same academic subjects across students. Measures of academic achievement like self-ratings, teacher ratings, parent ratings and grade point averages were excluded to avoid the confounding with problem behavior (Algozzine et al., 2011; Hinshaw, 1992). Externalizing behavior problems was defined per the most broadly accepted definition as behavior that is overtly hyperactive, aggressive, or oppositional (Achenbach, 1966; Achenbach, Ivanova, Rescorla, Turner, & Althoff, 2016). Additionally, given that Attention Deficit Hyperactivity Disorder (ADHD) was classified as a disruptive behavior disorder until recently (American Psychiatric Association, 2013), studies that examined children with ADHD were also included in this review.

Lastly, to enable potential causal interpretations and given the evidence of strong associations of IQ and language delay with both behavioral problems and academic achievement (e.g., Baker et al., 2007; Chow & Wehby, 2018; Hinshaw, 1992) methodological designs that did not account for this association were excluded from the
review. Research designs had to either: (a) include IQ or a language measure score as a covariate; (b) control for children who had been diagnosed with a speech, language, developmental or intellectual delay; or (c) explicitly state that they excluded children with either intellectual, developmental, speech or language delays from their sample. The latter choice was made primarily to include studies that may not have directly controlled for language or cognitive ability but would still allow potential valid inference about the relations between externalizing behavior and achievement. Given that the purpose of the research design was to reduce known potential confounds as much as possible, studies that did not directly control for language or intellectual ability, but excluded children with these disabilities from their sample would reduce the confounds of intellectual and language disability on the relations between externalizing behavior and academic achievement.

**Search Procedure**

Six databases were chosen for this review: OVID Medline, PsycInfo, Academic Search Premier, ERIC, Education Source, and PROQUEST databases for social science, health, and medicine in an effort to include dissertations and theses. The combination of search terms included terms for behavior, academic achievement, and statistical analyses to maximize yield of relevant articles given the broad domains of interest. There were a total of 18 search terms for behavior (e.g. “disruptive behavior*” OR “ADHD” OR “hyperactive*” OR “problem behavio*” OR “antisocial” OR “defian*”), 17 for academic achievement ( e.g. “academic fail*” OR “underachiev*” OR “academic achiev*” OR “school failure” OR “read* diffic*” OR “math diffic*”), and 11 relating to methodology (e.g. “path analy*” OR “developmen* cascade*” OR “longitudina*” OR “cohort stud*”)
OR “propensity scor*” OR “causa*”). For a complete list of search terms please refer to supplementary materials. This yielded 1,808 articles and 182 dissertations and theses. After duplicates were removed, 1226 articles remained for a title-abstract screen. After an initial abstract-title screen 388 articles were retained for a full text review. At this stage, a few articles frequently cited in the literature (e.g., Moilanen et al., 2010) were excluded as they did not meet the inclusion criteria for either IQ or language ability control. After full text review, 31 articles remained for a detailed methods appraisal. As a result, nine more articles were excluded. Most of these articles were excluded because they did not report discrete effects of externalizing behavior as defined in this review, e.g., the outcome measure included internalizing symptoms or other measures of behavior. An ancestral search of these 22 articles yielded 4 more relevant articles. Additionally, the journals Development and Psychopathology, and Developmental Psychology were hand-searched with the search terms “academic” and “externalizing.” No additional articles screened met the inclusion criteria, thus the final yield was 26 articles. See the PRISMA flow chart in Figure 2 for a more detailed description of the search and screening method.

**Coding and Quality Appraisal**

A quality appraisal was completed for these 26 articles based on Tooth, Ware, Bain, Purdie and Dobson’s (2005) *Quality of Reporting of Observational Longitudinal Research* and Hinshaw’s (1992) requirements for study characteristics that allowed for causal inferences between externalizing behavior and low achievement. The following characteristics were coded: (a) sampling strategy and sample frame (e.g., target population described, years of data collection, eligibility, selection in criteria, sample size); (b) sample descriptives provided for race, sex, age, socioeconomic status,
geographical location, etc.; (c) whether attrition rate and missing data were reported and accounted for; (d) reliability of measures (reliability of measures for the sample, scale, etc.); (e) if both variables of interest were measured at initial data collection and follow up; (f) child age at each wave of data collection; (g) theoretical confounds included in design; (h) statistical analysis (including reporting of power analysis, confidence intervals, effect sizes, standard errors and model estimation); (i) discussion of potential bias influencing results; and (j) whether the article addressed issues of external validity. If effect sizes or parameter estimates were not available in the full text of the article, authors were contacted in an effort to report them. In total, five author teams were contacted, and three provided effect sizes not present in the published literature.

Interrater reliability was calculated by dividing the number of items in agreement over the total number of ratings and converting the value to a percentage. This was completed at three levels: Title and abstract screen, full text review as well as coding and quality analysis. First, one primary coder coded all articles. Around 30% or all articles at all levels were randomly chosen and distributed evenly amongst three graduate student coders with the help of Microsoft Excel’s (version 16.16.3) random sort function. The rate of agreement was 99.2% with the primary coder for the title and abstract screen. For full text review, the rate of agreement for exclusion was 100% with the primary coder, and the rate of agreement was 94% for inclusion. Inter-rater agreement with the primary coder for coding of study features was 95.3% and the quality analysis agreement was 92.5%. All discrepancies in the coding process were reconciled using a collaborative and consensus-based process until an agreement was reached. Ultimately, 26 studies met criteria for inclusion. One of these, Duncan et al. (2007), was a combination 6 studies of
which two met inclusion criteria and thus these were included in the review.

**Results**

In total, 26 studies met criteria for inclusion. A total of 39,888 children were sampled across the 26 studies, with considerable variability in sample size (35 to 12,469), and demographics (see Table 1), indices of socioeconomic status, measures of standardized achievement and behavior (Table 2 and 3). Most samples were from the United States ($n=18$), others were from Canada ($n=3$), United Kingdom ($n=3$), or New Zealand ($n=2$). All studies were longitudinal in design. The quality appraisal coding for each study included in the review can be accessed via the open science forum, including a copy of the manuscript version of this study ([link redacted](#)). Then, we present evidence relating to each research question, starting with the evidence for causal associations relative to previously identified pathways, followed by consideration of potential moderating and mediating mechanisms identified from this empirical base.

To answer the first question for pathways, we first considered whether studies collected measures of achievement and externalizing behavior at both initial and follow up, thus allowing examination of bidirectional pathways or if they collected only one of each measure at initial or follow up, thus examining unidirectional pathways. The former design also accounts for the possible within time covariance of externalizing behavior problems and academic achievement, thus providing parameter coefficients between the time-ordered externalizing and achievement variables that contain less bias (Hinshaw, 1992). We end with synthesis and conclusions based on the methodological quality of studies ([summarized in Figure 3](#)) reviewed.
Sample Characteristics

A total of 39,888² children were included across 26 studies. There was considerable variability in sample size and demographics (Table 1). Generally, articles not from the United States did not report descriptives of race of their sample. Measures of SES used in analyses were heterogeneous. Measures reported included Hollingshead Index \((n = 2)\), parents’ education level \((n = 9)\), household annual income level \((n = 5)\), cumulative risk or adversity indices \((n = 3)\), percent of free and reduced lunch \((n = 2)\), or simply stating whether the sample was mostly middle, upper or lower class \((n = 2)\). Four studies did not report any measure of SES.

Further, even though all studies were longitudinal, there was still significant variability in study design. Six studies used group longitudinal designs, either dichotomizing externalizing behavior problem scores \((n = 4)\), reading scores \((n = 1)\) or both \((n = 2)\). There were also differences in sampling strategy; half of the studies \((n = 13)\) purposefully screened for children who were at risk for disruptive behavior difficulties or already had externalizing behavior problems, whereas the rest used datasets that were sampled from communities or schools with no emphasis on oversampling either externalizing behavior problems or low achievement. The duration of the studies varied as well. The mean gap between initial and final follow up assessments was 6.46 years, with the minimum being 2 years and the maximum being 11 years. Table 1.1 summarizes the descriptive properties of the studies. For detailed information of measures and covariates used in each study please refer to tables 1.2 and 1.3

² Datasets used by more than one study were only counted once and sample sizes are not added cumulatively in the descriptive statistics.
Does Evidence Support a Causal Relationship Between Externalizing Behavior Problems and Low Academic Achievement?

Overall, this review revealed little support for a causal relation between externalizing behavior problems and low achievement. The most common finding across studies was of non-significant effects and very small effect sizes, thus making conclusions about even predictive relations challenging. Below, we consider the respective support for bidirectional and unidirectional pathways.

Studies that examined bidirectional pathways. Six studies explicitly stated that they examined parallel and bidirectional effects (Table 2). Five started data collection when children were in kindergarten or earlier. Of these, two studies reported small effects for unidirectional pathways but in opposite directions, i.e., one from academic achievement in first grade to externalizing behavior in third grade (Burt & Roisman, 2010), and the other from externalizing behavior in kindergarten to academic achievement fourth grade (Hurry et al., 2016). There were studies that reported statistically significant bidirectional effects (e.g. Weeks et al., 2013) but the effect sizes derived from the standardized coefficients were all smaller than 0.1, discouraging meaningful interpretation (Cohen, 1994). The rest \( n = 2 \) reported neither significant nor meaningful effect sizes. Overall, there was contrary evidence presented, with no discernable trend supporting the existence of parallel effects.

Studies that examined unidirectional pathways. Studies reported in this section only examined unidirectional pathways and did not test parallel bidirectional pathways. The most common unidirectional pathway examined was that of early externalizing behavior problems to later achievement \( n = 19 \). As with analyses of bidirectional effects, the
predominant finding of this subset of the literature was of null or very small effects, including among studies that controlled for initial performance on the outcome. Few studies ($n=4$) found small to moderate effects of externalizing behavior problems to later lower academic achievement, and one found significant, but non-meaningful, effects from academic achievement to externalizing behavior (Table 2). Interestingly, a quasi-experimental study (Brennan et al., 2012) reported that early externalizing behavior predicted later academic achievement. However, when they examined the effects of an intervention to reduce externalizing behavior on students’ academic achievement, they found that only aggressive, not hyperactive, behavior had a significant effect on academic achievement ($\beta = 95\% \text{ CI } [0.002–0.293]$). The rest of the studies ($n=14$) found either nonsignificant or non-practical effect sizes.

This inconsistency, or lack of support, also characterized results across developmental periods. For example, studies that collected measures in preschool- which strengthens temporal precedence assumptions- through elementary school, also provided only tenuous evidence, even where similar statistical designs, covariates and sample frames were utilized. For example, despite similarities, Brennan et al. (2012) found small effects between early externalizing behavior and later academic achievement ($\beta = -0.16$ to -0.14, $p < .01$), whereas Gray et al. (2014) found no effects. There were no trends between teacher or parent reported externalizing behavior and relations to academic achievement or vice versa. Overall, more than half of the studies reported non-significant

3 Though we were able to calculate a small increase in hyperactivity scores ($\Delta d = 0.2$) in the reading group compared to the control over time in one study (Willcutt et al., 2001), the small sample size and non-significant t tests precluded support of the reverse pathway between early academic achievement and later externalizing behavior.
effects, and even among those that examined similar developmental spans and had similar research designs, contrary results were reported.

**What Are the Mechanisms Moderating or Mediating the Relationship?**

As with the first research question, there was little empirical support from which to draw conclusions about potential mediators and moderators. As this research question dealt with identifying variables that potentially influenced the relation between academic achievement and externalizing behavior problems, structural equation modeling techniques were especially suited for these kinds of hypotheses (Kline, 2015). Of the nine studies that used this method, only two (Burt & Roisman, 2010, Weeks et al., 2013) examined potential mediating effects on the relations between externalizing behavior and academic achievement. That said, authors did not explicitly test these cross-domain pathways for significance, instead only reported indirect effects, rendering conclusions about mediation challenging (Kline, 2015). The studies that addressed moderating variables explicitly in their analysis found no meaningful differences in variation by gender (Burt & Roisman, 2010) or race (Kremer et al., 2016). However, there was one study that found significant negative interactions of executive functions and ADHD symptomology on academic achievement (Rennie et al., 2010; $\beta = -0.45, p < .01$).

**Covariates.** Lastly, half of all included studies directly controlled for IQ ($n = 13$); some controlled for early language ability ($n = 5$), whereas two studies controlled for children who were diagnosed with speech or intellectual delays. The remainder ($n = 6$) excluded children with speech/language, intellectual and developmental delays from their sample. Overall, results were similar and there was no difference in the presence of significant effects based on how studies accounted for language or intellectual ability.
Additionally, all studies that directly controlled for language and IQ found moderate relations of these controls to academic achievement and behavior (Table 2). There was a great diversity in additional covariates added to studies. However, covariates included in research designs could fit two primary categories; (a) inattention and executive functioning measures, or (b) social skills and internalizing measures based on theories of developmental cascades (e.g., dual failure models). The most common covariates found associated with academic achievement were inattention, socio-economic measures, as well as early measures of academic achievement. When the outcome was externalizing behaviors, the most common covariates found to be significant were sex, socio-economic status, and early measures of externalizing behavior (Table 2).

**Summary and Quality Analysis**

Though lower quality studies tended to report larger effects when present, overall, there was no discernable association of study quality to the direction or existence of a relation. For example, two of the five highest quality studies found small to very small significant effects, but of opposite directions, and two of the five lowest quality studies reported significant small to medium effects also of opposite directions (Table 1.4). According to Sackett, Straus, Richardson, Rosenberg, and Haynes’s (2000) best evidence synthesis suggestions, findings can be described as consistent when 75% of the studies agree on the existence and direction of a specific relation. Of the 21 studies that tested the direction of the relation from early externalizing to later achievement, 43% indicated statistically significant relations. Of these, if we only consider studies that had meaningful effect sizes, then only 23% supported the existence of a relation. Similarly, of the 10 studies that tested the direction of the relation from achievement to later
externalizing behavior only 30% of the studies found significant effects, of which only two found small, meaningful effect sizes. The sample size of the study did not seem to have an association with study quality or the reporting of significant results (Table 2). Though we were able to report small effect sizes for two of the smaller scale studies, these accompanied non-significant statistical results, likely caused by underpowered analyses, and thus discouraging meaningful interpretation of effect sizes (Schäfer & Schwarz, 2019).

Studies that did not model cross-domain covariance at initial or follow up, thus introducing potential bias in their estimates (Hinshaw, 1992), did not find significant results at a rate lower or higher than studies who did model cross-domain covariance. Kraemer et al. (2001) suggested that, compared to dichotomous variables, variables on a continuous scale might overestimate parameter estimates between two variables. Results do not support this hypothesis, however, given the unequal proportion of studies in each type of design (Supplementary tables) conclusions about generalizability should be drawn with caution. Thus, taken together, evidence to validate potential causal or even predictive associations between achievement and behavior strongly favors the hypothesis that they are not directly related, especially when taking covariates into account.

**Discussion**

The present study systematically reviewed current literature on the relationship between externalizing behavior problems and academic achievement in an effort to update Hinshaw’s (1992) summary on potential causal mechanisms between the two domains. Overall, results suggested that there was little evidence for a causal association- or even a predictive one- between externalizing behavior problems and academic
achievement, or even a predictive association. No study found evidence for bidirectional effects and only five studies found evidence for unidirectional effects, most commonly for early externalizing to later lower achievement. Thus, results were consistent with Hinshaw’s (1992) review as there was inconclusive evidence of either unidirectional or bidirectional causal pathways; however, unlike Hinshaw’s (1992) review, this study did not find any evidence for a predictive relationship between the domains. A primary reason for the non-significant effects could be that most studies reported strong associations of antecedent variables with their outcomes, thus potentially accounting for the relationship between externalizing behavior problems and achievement (Hinshaw, 1992; Algozzine et al., 2011). In particular, and similar to previous research (e.g. Duckworth, Quinn, & Tsukayama, 2012; Murray & Farrington, 2010), cognitive ability was the most consistent antecedent variable to both lower achievement and higher externalizing behavior problems (Table 1.4), though effect sizes were larger when the outcome was academic achievement. Notwithstanding, the extent to which these findings can be generalized is limited due to the wide variability in the methodologies of the studies reviewed.

Similarly, expressive language was an important antecedent when the outcome was achievement, as all studies that controlled for early expressive language found small to moderate relations to academic achievement. This is in line with current theory and empirical evidence that acknowledges the critical role language plays in both reading and math development (Peng et al., 2020). Similarly, a recent longitudinal review reported small but consistent correlations between early expressive language and behavior and suggested that there could be several potential variables mediating the relation (Chow,
Ekholm & Coleman, 2018). There were few studies that controlled for language when the outcome was externalizing behavior \((n = 2)\), and studies reported contradictory results—i.e., one found language had moderate effects on behavior (Weeks et al., 2013), whereas the other did not (Fitzpatrick & Pagani, 2013). Given the small sample size, it is difficult to generalize the impact early language has on the relations between externalizing behavior and achievement based on this study.

Though covariates included in the analyses were heterogeneous, another consistent finding was that studies that included inattention as a covariate were more likely to find non-significant effects of externalizing behavior problems on later achievement outcomes. This supports claims that inattention, not aggressive or disruptive behavior, is associated with future lower achievement (e.g., Fergusson & Horwood, 1995; Gray, Duck, Rogers & Tannock, 2017). Inattention was also observed to be a consistent predictor of externalizing behavior, suggesting that it plays not only an important role in academic achievement, but also the development of externalizing behavior (e.g., Grimm et al., 2010). As such, an important implication for future research is that studies that do not include inattention as a distinct variable could be missing an important confounder in their analysis.

**Heterogeneity in Research Design**

Studies varied in measures, covariates, and statistical design. A majority of the articles in this review utilized measures identified in a recent meta-analysis as gold standard for assessing externalizing behavior (Achenbach et al., 2016); however, the type of rater (e.g., parent or teacher) varied across studies (Table 2 and 3). Even though recent meta-analyses report that parents and teachers have higher rates of agreement for
externalizing behavior problems than internalizing behavior problems (De los Rayes et al., 2015), and there was no discernable pattern of rater and statistical result, there is conflicting evidence on the validity of teacher ratings of problem behavior, especially across different student groups (e.g., Grietens et al., 2004; Mason, Gunersel & May, 2014). Additionally, no study in the paper included direct observations of behavior, and there is evidence suggesting direct observations could provide additional validity evidence of externalizing behavior, as opposed to relying only on rating scales (Clemens, Shapiro & Siebert, 2016). Thus, the manner in which researchers chose to measure behavior may have influenced results.

Unfortunately, similar covariates may not be enough to ensure replicability and generalizability of results. Studies also had considerable heterogeneity in the statistical methods used to answer research questions (Table 2). Even when the same dataset was analyzed (e.g., Burt & Roisman, 2010; Duncan et al., 2007), researchers chose different statistical methods and reached different conclusions (see Grace & Bollen, 2005 for a detailed review on SEM versus multiple regression). This methodological diversity hinders quantitative summarization able to answer causal relations, given that the beta estimates provided in each study are interpretable only in the context of the study itself (e.g., Greenland, Maclure, Schlesselman, Poole & Morgenstern, 1991). However, the existing evidence, regardless of method, covariates chosen, and sample, does not support a valid or reliable predictive or causal relation between achievement and behavior. Future researchers interested in examining the potential causal relations between the two domains could choose to replicate designs of previous high-quality articles in an effort to validate findings.
Implications for Practice

There is a widely-held assumption that externalizing behavior problems and academic achievement are related, yet the results of this review provide little evidence that externalizing behavior problems and academic achievement are causally related (e.g., Algozzine et al., 2011), or even temporally related (e.g., Duncan et al., 2007, Metsäpelto et al., 2015). The paucity of evidence could be challenging for those who believe that intervention in one domain positively influences performance in the other. Algozzine et al. (2011) proposed that the solution to this tension is to continue to concentrate on “what we know” (p.13); that is, teaching both academics and behavior.

Thus, 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 (Lane, Oakes and Menzies, 2014). In other words, the prevention of academic and behavioral problems should depend on evidence-based instruction or intervention in academics and behavior instead of expecting gains in one domain to carry over to the other (Corcoran, Cheung, Kim, & Xie, 2018; Sugai & Horner, 2009). This implication applies across early childhood education and K12 education, with teaching of both academic and behavioral skills important to school readiness (Biermen & Motamedi, 2015; Schindler et al., 2015). Lastly, results of this study imply that antecedent variables like inattention, language, cognitive ability, parenting, and family SES are likely associated with both behavior outcomes and academic achievement. In practice, this lends support to the utility of assessing language skills in children diagnosed with both behavioral and academic difficulties in order to improve targeted intervention (e.g., Hollo, Wehby, Oliver, 2014; Peng et al., 2020). It also underscores the importance of policies that grant not only
access to free, high-quality early child care programs, but also wraparound services intended to support the well-being of the entire family (Dodge & Haskins, 2015).

**Limitations and Future Directions**

Although this study can help clarify persistent debate about the relations of externalizing behavior and achievement, it is not without limitations. Firstly, this study only included articles written in English (Ganann, Ciliska & Thomas, 2010). Second, a possible limitation and an area for future research is that this review only included studies that controlled for IQ or language ability in an effort to eliminate alternate explanations and strengthen causal hypotheses; however, future reviews could include studies that did not control for these variables, thus enabling a comparison.

Lastly, even though this review addresses causal relationships, it is important to note that without an actual dose-response manipulation, it is difficult to validate a causal theory (Kazdin, 2007). However, this does not eliminate the possibility of building an argument for a causal association between two variables (Kraemer et al., 2001). Additionally, well designed mediation and moderation studies are needed to address why two domains are, or are not, related and whether the association is stronger for certain groups of students, e.g., clinical populations (Cole & Maxwell, 2003; Kraemer, Kiernan, Essex, & Kupfer, 2008). Specifically, there is evidence to suggest that externalizing behavior does not develop in a homogenous way (Moffit, 1993), which was a common assumption in all the studies in the sample. Thus, even though there may not be a need for additional correlational or multiple regression studies as Algozzine et al. (2011) suggest, addressing these gaps in the literature could help build more conclusive evidence on the relationship between externalizing behavior problems and low academic achievement.
Chapter 3: Early Academic Achievement and Relations to Trajectories of Externalizing Problem Behavior in Elementary School

The notion that P12 students’ academic and externalizing behavior outcomes are linked is so ingrained in the minds of educators (Algozzine, Wang & Violette, 2011) that not only is it the basis of several interventions (e.g., Jones, Brown & Aber, 2011) but public policy, too (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, 2013), and school-wide positive behavioral supports (SWPBIS; Horner & Sugai, 2015) to improve not only students’ behavioral outcomes, but academic ones as well. Unfortunately, research on the causal association between externalizing behavior and academic achievement is limited, if not tenuous (e.g., Duncan et al., 2007; Algozzine et al., 2011). Yet evidence on the causal and even predictive association between externalizing behavior and academic achievement is mixed at best, if not tenuous (e.g., Kulkarni et al., 2020; Study 1).

Further, the common assumption that the association is the same within the entire student population could mask differential group mechanisms (Kazdin, 2007). Given that current empirical evidence identifies at least three different group trajectories of externalizing behavior (e.g. Olsen, Choe & Sameroff, 2017), considerations of how associations could differ across trajectories of externalizing behavior could support more efficient and effective targeting of scholarship and professional resources. Lastly, based on the gaps in the literature identified in study 1, re-examining the relationship between the two domains, albeit with appropriate controls and a rigorous research design, could
not only help clarify the relationship between externalizing behavior and academic achievement, but provide evidence that intervening in one domain could indeed prevent development of difficulties in the other. The primary purpose of this study, therefore, is to add to the literature base on the association between academic achievement and externalizing behavior problems by using 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.

**Brief Review 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 drop out (Piquero, Jennings, & Barnes, 2012). Externalizing problem behavior was thought to remain largely stable through elementary school (e.g. Campbell, Spieker, Burchinal, Poe and the NICHD ECCRN, 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 in the course of elementary school (e.g., Olson, Choe & Sameroff, 2017). Several variables such as teacher student relationships (Olson et al., 2017), race (Miner & Clarke-Stewart, 2008), and parental attachment (Feldman, Masyn & Conger, 2009) are thought to have an association with trajectory membership.
Despite a large body of research examining the relationship between externalizing behavior and academic achievement (Hinshaw, 1992; Study 1), most studies have neglected to consider that the association between the domains could differ amongst subpopulations of students who display heterogenous development of problem behavior. This is especially important given that the assumption of homogeneity within the population could result in inaccurate conclusions about associations between two domains (Kazdin, 2007). The ability to validate an association between the two could be critical for early intervention, as intervening on a causal mechanism when children are younger could have cascading impacts on longitudinal outcomes that are related to the deficits (Cicchetti & Masten, 2010).

**Theories of association.** Though current evidence on the predictive association is mixed (Algozzine et al., 2011), studies continue to report moderate to high levels of comorbidity between low achievement and externalizing behavior (e.g. Wagner, Kutash, Duchnowski, & Epstein, 2004; Tannock et al., 2018). The development of comorbid behavioral and academic problems is predominantly conceptualized within the dynamic systems theory, which suggests that changes in one area of functioning can trigger a sequence of consequences that ultimately have broader developmental effects (Sameroff, 2000), or a developmental cascade (Cicchetti & Masten, 2010). For example, if a child enters school with significant disruptive behavior, the resulting poor interactions with peers and social rejection could reinforce maladaptive attention seeking behavior, ultimately leading to less opportunities to experience academic success (Patterson, DeBaryshe & Ramsey, 2017). These children are also more likely to face a higher number of school disciplinary actions including suspension and expulsion, resulting in
reduced instruction time, and consequently low academic achievement (McIntosh, Horner, Chard, Boland & Good, 2006).

Academic failure could also be a trigger to comorbid difficulties (Metsäpelto et al., 2015). Children who expect failure in academic tasks may learn that engaging in maladaptive behaviors, commonly disruption, allows them to escape the undesired academic task (McIntosh et al., 2006). This cycle of behavior is often reinforced in school settings by disciplinary consequences that enable the escape motivated behavior (e.g., time out, suspensions). This dual failure in behavior and academics is thought to exert cross-domain effects as students complete their school career and can lead to higher probabilities of adult psychopathology (Masten et al., 2005) and even incarceration (Piquero et al., 2012).

**Evidence of associations.** The last systematic review conducted on the topic of causal associations between externalizing behavior and academic achievement suggested that there were small to moderate associations between the two domains, however Hinshaw’s (1992) own conclusions were overwhelmingly cautionary given the lack of studies with appropriate methodology. Similar findings are reported in Study 1 and currently, published evidence on a valid association remains limited (Algozzine et al., 2011). For example, some researchers have reported small to moderate negative effects in the direction of early behavioral difficulties and lower academic achievement in middle and high school (e.g., Masten et al., 2005; vanLier et al., 2012). A smaller literature base proposes that significant negative associations exist in the opposite direction, i.e., early academic difficulties are associated with later externalizing behavior (e.g., Burt & Roisman, 2010). However, many of these findings are contradicted by other studies who
have found no significant associations between the two domains (e.g. Duncan et al., 2007; Metcalfe, Harvey & Laws, 2013; Morgan et al., 2018).

At the school level, a recent meta-analysis concluded that the most rigorous randomized control trials of social-emotional learning implementation showed no improvement in academic outcomes (Corcoran, et al., 2018) and SWPBIS has also shown mixed effects on academic gain (Noltemeyer, Palmer, James & Wiechman, 2018). Differential conclusions across research designs have been theorized to be, in part, due to heterogeneous study samples, measurement error, and inadequate statistical controls (Grimm, Steele, Mashburn, Burchinal, & Pianta, 2010; Hinshaw, 1992). For example, inattention (e.g. Metcalfe, Harvey & Laws, 2013; Duncan et al., 2007), early language skills (Chow & Wehby, 2018) as well as socioeconomic status (Lam, 2014) are associated with both achievement and behavior, and thus studies that do not account for these confounds could report inaccurate estimates (Hinshaw, 1992). Other results have indicated that gender and race could have a moderating effect on the association between the two domains (e.g. Kremer, Flower, Huang & Vaughn, 2017).

**Research Purpose**

Given the challenges in building an evidence base for causality with psychological constructs (Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007), it is not surprising that despite several studies examining the association between externalizing behavior and academic achievement, conclusive evidence remains limited (Algozzine et al., 2011). However, the ability to validate causal associations is integral to scientific fields interested in improving outcomes and evaluating the effectiveness of interventions (Glass, Goodman, Hernen & Semet, 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). Study 1 attempted to review the literature in an effort to reveal potential causal mechanisms and I found several methodological and theoretical drawbacks in the literature including inadequate controls, heterogenous methodological designs, and the assumption that children’s externalizing behavior develops in a homogenous manner. Study 2 will attempt to fill some of these gaps in an effort to add to the literature on the association between achievement and externalizing behavior to support conclusive evidence on potential causal or predictive relations.

This study was, therefore, motivated by three research needs. Firstly, intervention implementation—be it small group, class-wide or systems-wide—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 intervention efforts and potential outcomes, in this case, whether intervening on academic achievement, can indeed result in improved behavioral outcomes or vice versa (Bradshaw, Zmuda, Kellam & Ialongo, 2009). Secondly, identifying whether the relationship is stronger for certain subgroups of students is critical to inform prevention efforts, especially in kindergarten and first grade, as schools could target intervention efforts on the mechanism to improve both academic and behavioral functioning (Masten & Cicchetti, 2010), and broader outcomes shown to be related to deficits (vanLier et al., 2012). Lastly, the history of contrary findings on this issue contributes to the replication crisis in psychology (Maxwell, Lau & Howard, 2015).
and this study will add to the body of literature, leveraging a nationally representative longitudinal sample, appropriate controls, and a rigorous research design.

Specifically, the research questions to be answered are:

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?

Method

Data Sample

This study will utilize the Early Childhood Longitudinal Study, Kindergarten cohort of 2011 (ECLS-K:2011) collected by the US Department of Education’s National Center for Education Statistics. The Early Childhood Longitudinal Study has several advantages over other nationally representative datasets. It not only is the most current and therefore representative of the current cohort of elementary students in the United States, it is 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. Additionally, to enable reliable statistical analysis, Asian, Native Hawaiians, and Other Pacific Islander children were oversampled. A multistage, clustered sampling design was used 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 fall kindergarten. Ten of these were nationally representative, whereas 20 were non-
representative and selected by stratified random sampling of the remaining PSUs based on geographic region, size, socioeconomic status, race/ethnic distribution, and metropolitan areas. The final sample of students (N= 18,170) was followed up to fifth grade unless the child died or left the United States. This study will use child level data collected in the springs of kindergarten (2010–2011), first, second, third, fourth, and fifth grades. To account for attrition, missingness, and the clustered nature of the data, the authors provided several weights and their corresponding PSU and strata for researchers to use.

**Measures**

Teacher, instead of parent ratings were used to construct behavioral measures because the behaviors of interest were in the school setting. This decision is supported by the low to moderate agreement between parent and teacher ratings of behavior across home and school settings (De Los Reyes et al., 2015).

**Externalizing Behavior.** The externalizing behavior subscale ($\alpha = 0.87$ to 0.89) for spring kindergarten, first, second, third, fourth and fifth grade will be used as the longitudinal outcome measure in this study. It is from a modified version of the Social Survey Rating System (SSRS; Gresham & Elliot, 1990) designed for ECLS. Items included statements about if the child disobeys rules, fights with others, has temper tantrums, and so forth. The variable presented in the dataset is the mean of the summed of scores on this scale (4 = almost always to 1 = Never). Kindergarten and first grade spring subscales had five items whereas the rest had six.

**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. Outliers were flagged and will be removed for analysis. Reading and math achievement in spring kindergarten ($\alpha = 0.95, 0.94$) and first grade ($\alpha = 0.93, 0.93$) will be 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., 2015).

**Behavioral covariates.** Teacher reported measures of approaches to learning and inattention in spring of kindergarten will be used as covariates given previous research suggesting that both could act as confounds of the relationship between externalizing behavior and achievement (Duncan et al. 2007, Gray et al., 2014). The approaches to learning measure was developed for the ECLS based on the dimensions of school readiness described by the National Education Goals Panel framework (NEGP; Kagan, Moore, & Bradekamp, 1995) to encapsulate classroom-based behaviors considered conducive to learning in a school environment. The measure included six items ($\alpha = 0.91$) that asked teachers to report on how often students exhibited a selected set of positive learning behaviors based on the NEGP definition. For example, 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 subscale of the Children’s Behavior Questionnaire in the spring of kindergarten (CBQ; Putnam & Rogers, 2006). The CBQ asked teachers to score how often children exhibited behavior on a 7-point scale (1=extremely untrue to 7 =extremely true). The inattention subscale of the CBQ has been validated for both racially diverse and low-income samples (Putnam & Rogers, 2006; Richard, Winders & Burns, 2008).
Cognitive and language covariates. Given that cognitive and early language abilities have moderate associations with externalizing behavior and achievement (e.g., Rennie, Beebe-Frankenberger & Swanson, 2014; Chow & Wehby, 2018), both will be used as covariates. The model will include a direct cognitive assessment of working memory where children are asked to complete a backward-digits task that requires the child to orally repeat numbers in the reverse order that was presented to them (Woodcock-Johnson Psychoeducational Battery, 3rd edition). Additionally, for early language skills, the ECLS includes scores two tasks from the Preschool Language Scale (preLas 2000). The “Simon Says” task required students to follow simple spoken instructions in English and the “Art Show” task was a picture vocabulary assessment (Tourangeaue et al., 2015). The combined score of the language screener in the spring of kindergarten will be used in the analysis (α = 0.89). These were measured in the fall and thus can be considered antecedent variables in the models.

Socio-demographic covariates. This analysis will also control for sex, race, family socioeconomic status, parent’s education attainment, and percentage of non-white students in school in an effort to account for associations between socio-demographic indicators (e.g., Lam, 2014; Morgan et al., 2018) on externalizing problem behavior and academic achievement.

Data Analysis

The first research question seeks to answer if students’ early academic achievement is associated with membership to externalizing behavior trajectories in elementary school. Secondly, what child and family characteristics are associated with membership to these trajectories. The ECLS authors 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 (O’Connor, Dearing & Collins, 2011; Shiyko, Li, & Rindskopf, 2012). Thus, to account for the original frequency characteristic of the measure and the non-normal positive skew, the mean will be multiplied by the number of items on the subscale. Though the kindergarten and first grade subscales had one less item (five, not six), all means will be multiplied by six to maintain comparisons with other rounds of data collection. Thus, the total sum frequency of behaviors for each student will be modeled (O’Connor et al., 2011) instead of 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 will be subtracted by 6, shifting the range of frequencies from 6, 24 to 0, 18. As a result, children who did not show any externalizing behavior in any item on the scale will have a score of 0, rather than 6, allowing clearer interpretation without disturbing the properties of the original distribution. For each wave the outcome measure was modeled as a bar plot to examine the distribution of the counts (Figure 1). The distribution will inform the model used and is discussed in more detail below.

**Missing Data.** Using the weight to select cases to be in the analysis, it was found that less than 5% of the data were missing across every variable and time point except for birth weight and IEP status for which around 10% and 8% of data was missing. Additionally, given that ECLS consists of several related variables including parent and teacher reports of inattention, internalizing, approaches to learning, gender, sex,
socioeconomic status, birthweight, this dataset qualifies for the MAR assumption that allows for Full Information Maximum Likelihood Estimation (Bhaskaran & Smith, 2014), which is a robust method to estimate missing data in structural equation modeling methods even with non-normal data (Larson, 2011). However, before delving into the statistical analysis, researchers must choose the appropriate distribution that best matches their data (Greenwood et al., 2019).

**Growth Curve Modeling and Mixture Modeling.** Observed heterogeneity in traditional growth curve models can be explained by covariates that account for variation in intercept and slope (Reinecke & Seddig, 2004). However, in cases where individuals are theorized not to come from the same population, mixture modeling provides researchers a way to model unobserved heterogeneity (Reinecke et al., 2014). It can be seen as an extension of latent class analyses, where class membership is assigned to each developmental trajectory (Nagin, Jones, Passos, & Tremblay, 2018). Within the scope of mixture modeling, Nagin & Tremblay (1999) introduced what was called a group-based approach, or a latent class growth approach (LCGA) which assumes units in one class have a homogenous intercept and slope, i.e. within class homogeneity of variance around the intercept and slope is assumed. They used this approach primarily to model heterogenous developmental trajectories of anti-social behaviors within the population (see Nagin, 2005). Muthén (2004) consider the LCGA to be a simplified specification of a growth mixture model (GMM). The 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” (page 417, Reinecke & Seddig, 2004).
In this study, the latter method was chosen due to (a) the purpose of this study being better suited to the assumptions required of LCGA, i.e. identification of latent trajectories of problem behavior and their association with academic achievement and other child level factors (Nagin et al., 2017; June & Wickrama, 2008) and (b) the complexity of GMM often yielding unreliable solutions due to nonconvergence of multiple local maxima when modeling behavior (e.g. Feldman, Masyn, & Conger, 2009; Olson, Choe, & Sameroff, 2017).

Most literature using latent class growth analyses has modeled academic trajectories (e.g. reading) and examined child, family and school characteristics related to various trajectories of academic achievement (e.g. Boscardin, Muthén, Francis, & Baker, 2008). Fewer studies have examined trajectories of externalizing problem behavior, however, several have established at least three trajectories of problem behavior, providing evidence that subgroups of behavioral growth exist within the student population (e.g. Moffit et al., 1999; Olson et al., 2017). Never the less, the association of students’ early academic achievement with these trajectories is still an area of research.

In a latent class growth model, let \( Y_{it} (y_{i1}, y_{i2}, \ldots y_{it}) \) be the outcome variable for individual \( i \) at time \( t \). Then we can say that the conditional distribution of the observed data for student \( i \) given \( k \) classes and baseline covariates \( X (X_i=x_{i1}, x_{i2}, \ldots x_{in}) \) is

\[
f (y_{i}; x_i) = \sum_{k=1}^{k} \Pr (C_i = k; X_i = x_i) \Pr (Y_i = y_i; C_i = k) \ldots (1)
\]

In simpler words, the above equation, derived from Jones, Nagin & Roeder (2001) represents the likelihood of observing the data trajectory of a student, given his or her covariate values. Thus, this LCGA model will include the latent categorical variable \( C_i \), which represents the different possible trajectory subgroups, that is conditioned on
both the observed outcome variable (Y) and baseline covariates or risk factors (X). Figure 1.3 represents this model in a simple way. Here risk factors or base line covariates are considered time invariant because they were collected only once and are assumed to be associated with class membership and provide no other information to the model after accounting for class membership (Nagin & Tremblay, 1999). Statistically, this relationship is defined by a multinomial logistic regression for unordered responses where the log odds of being a member or particular class k versus being in the reference class K are modeled as a function of the time invariant baseline covariates X_i (Shiyko et al., 2012). Thus, in a latent class growth model, the latent class categorical variables follow a multinomial distribution in which the parameters \( \pi_{ik} \) give the probability of the underlying trajectory class \( k \) where \( k \geq 1 \) (Yang, 2015).

\[
\pi_{ik} = \frac{e^{z_i^T \gamma_k}}{\sum_{h=1}^{K} e^{z_i^T \gamma_h}} \quad \ldots (2)
\]

**Modeling Count Variables.** When modeling count variables, there are primarily two statistical options that are available to the researcher: the Poisson distribution and the negative binomial distribution (Allison, 2012). The Poisson model assumes that the conditional variance of the dependent variable is equal to the conditional mean (equi-dispersion), whereas the negative binomial model does not hold this assumption and is appropriate for both over dispersion and excess zeroes (Allison, 2012; Zeilies et al., 2008). Often, in behavioral research the assumption of equi-dispersion does not hold and over dispersion occurs, i.e. the variance is much larger than the mean, resulting in poorly fit Poisson models (Zeilies et al., 2008). To account for this poor fit, a Zero Inflated Poisson (ZIP) model is thought be another way to improve over-dispersion by accounting for data that has an overabundance of zeros (Lambert, 1992).
When modeling problem behavior, as in this study, an important factor to consider is that data often does include a large number of zeros due to many children showing low frequencies of externalizing behavior (Nagin & Tremblay, 1999). I.e. many children have no reported externalizing behavior problems. In previous seminal literature modeling count data using latent class growth approaches, most studies have used a ZIP model (e.g. Nagin et al. 2018). Conceptually, a zero inflated model theorizes that there are two types of zeros present in the data. One is structural, i.e. there is a group of children who naturally are not at risk for any externalizing behavior problems, and the other is random, for the group that is susceptible to externalizing problem behaviors but may still produce a zero-outcome due to sampling variability (Hua, 2014).

Currently, developmental theories identify a group of students with low externalizing behavior problems, and unlike other behaviors like smoking, externalizing behavior itself is a naturally occurring behavior in children (e.g. Olson et al., 2017). In other words, only elevated or clinically significant levels constitute problem behavior. Thus, there is no theoretical reason to use a zero-inflated model other than the higher frequencies of zeros in the data set. Allison (2012) recommends that a standard negative binomial model, which naturally accounts for over-dispersion and excess zeros, may work equally well and better than a ZIP model. Thus, this study will also run the first unconditional base model with a negative binomial model to test which model returns the best fit statistics. To do this, this study will compare model fit using the BIC values as criterion for the model with the best fit (Yang, Harlow, Puggioni & Redding, 2017). This is especially pertinent in this case as bar plots (Figure 2.1) show that not all waves of data had extreme zero inflation. The model with best fit will then be used to specify the
distribution to run the latent class growth analysis on. Below, a brief explanation of a standard ZIP follows.

**Zero Inflated Poisson Model.** A zero inflated Poisson model (ZIP) estimates two mixture models simultaneously. It combines the Poisson regression model with a logit model to account for the non-random zeros of the group that has naturally occurring non-event behavior (Lambert, 1992). Thus, in a zero inflated model, two equations are estimated to obtain desired probabilities. The first equation represents the logit model that accounts for the zero inflation, whereas the second represents the traditional Poisson model. Thus,

\[
\Pr(Y_i = y_{it}; C_i = k) = p_{it} + (1 - p_{it}) \frac{1}{e^{\mu_{it}}}, \text{ for } y_{it} = 0
\]

and

\[
(1 - p_{it}) \frac{\mu_{it}^{y_{it}}}{y_{it}! e^{\mu_{it}}}, \text{ for } y_{it} \geq 1 \ (1, 2, 3\ldots) \quad (3)
\]

Here, \( p_{it} \) is the probability of an individual having a non-random frequency of zero, i.e. being in the group that is naturally observed to have zero frequencies of problem behavior, and \( \mu_{it} \) is the parameter representing the mean frequency count of behavior for student \( i \) at time \( t \), given class \( k \). The notation of the parameter \( \mu_{itk} \) can be be can be summarized as follows when the rate of change is modeled as a linear function (Yang, 2015; Shiyko et al., 2012; Nagin & Tremblay, 1999). Higher order polynomials can be added into the model depending on the form.

\[
\log(\mu_{itk}) = \beta_{0k} + grade_{it} \beta_{1k}
\]

\[
\log(\frac{p_{itk}}{1 - p_{itk}}) = \alpha_{0k} + grade_{it} \alpha_{1k}
\]
Where $\alpha_{0k}$ and $\beta_{0k}$ are the intercepts for class $k$, and $\beta_{1k}$, $\alpha_{1k}$ are the slopes for class $k$. In the ZIP model, two regression coefficients are being modelled due to there being two models estimated simultaneously. Thus, the equation above denotes that the probability of a teacher observing externalizing behavior in grade $t$ would depend on the average observed rate $\mu$ for a particular grade $t$ for individuals in class $C_i = k$, given baseline covariates (Shiyko et al., 2012). When interpreting parameters each class will have two estimates, one for the structural zeros and the other for the regular count model. Random effects or within class variability in the growth model will not be estimated as within class variability is fixed in latent class growth models (Shiyko et al., 2012).

**Negative Binomial Models.** A negative binomial distribution (or Poisson-Gamma) is derived from a Poisson distribution but has an additional over-dispersion parameter $\phi$ (Reinecke, 2011). Overdispersion, in simple terms, is the presence the greater variability observed in the outcome than would be expected from a standard (Poisson) model (Allison, 2012). Poisson models assume that the variance is equal to the mean, but in the behavioral sciences, this is almost never the case and over dispersion is often the norm (Allison, 2012). Ignoring this characteristic of the data can lead to biased standard errors (Reinecke, 2011). A simple negative binomial model is written similar to the standard Poisson model, but with an added $\phi$ parameter to model overdispersion (Yang, 2015), i.e.

$$
\Pr(Y_{it}=y_{it}; C_i=k) = (1-p_{it}) \frac{\Gamma(\phi+y_{it})}{y_{it}!\Gamma(\phi)} \left( \frac{\mu_{it}}{\mu_{it}+\phi} \right)^{\mu_{it}} \left( \frac{\phi}{\mu_{it}+\phi} \right)^{\phi}, \text{ for } y_{it} \geq 1 \ (1, 2, 3...) \quad \ldots (4)
$$

Here, $p_{it}$ is the probability of an individual having a non-random frequency of zero, and $\mu_{it}$ is the parameter representing the mean frequency count of behavior for student $i$ at time $t$, given class $k$. Larger parameter estimates of $\phi$ indicate the presence of more
dispersion. If the negative binomial (NB) model is found to be a better fit for the data, then interpretation will be less complex. This is because NB models do not have two parameter estimates for each class as in a zero-inflated model and simple odds ratios can be reported to describe the likelihood of a student belonging to a particular class.

**Model Building.** Model building in mixture models is still an area of active research (Kamata et al., 2018). There are three main approaches to modeling latent class growth models: the classify-analyze approach, the one step or direct approach, and the three – step approach. Previously, the classify-analyze approach was used frequently in the behavioral sciences (Vermunt, 2010). In this approach, there are three sequential stages to estimate the model. First, the model without the covariates of interest is estimated to identify the appropriate number of classes. Second, this mixture model is used to derive and approximate class membership for each person in the data (e.g. people are assigned to the class that they have the highest probability of belonging in). In the last and third step, a separate multinomial regression is carried out with the variables of interest predicting membership into classes derived previously. Since class membership is decided using posterior probabilities, there is no ‘true’ or perfect class assignment, but this error is ignored when conducting the the multinomial regression. i.e. membership to class (or trajectory) is assumed to be without any error (Dziak et al., 2016). This approach has been criticized to have several methodological drawbacks due to ignoring classification error in assigning cases to classes, and is generally not recommended anymore (Kamata et al., 2018).

The second approach is the three-step approach introduced by Vermunt (2010). The first step in this method is similar to the classify-analyze method. However, in the
second step it corrects for bias due by retaining the information about classification uncertainties and utilizing it as the measurement errors of classifications. In the third step of this approach, the covariates are fit separately for each of the identified classes by incorporating the measurement errors derived in the second step (Vermunt, 2010). Recent simulation studies have shown that this method is most recommended for latent class analyses (e.g., Dziak et al., 2015; Kamata et al., 2018).

The third approach, and the approach used in this analysis, is called the one step or direct method, in which covariates are used simultaneously to estimate class trajectories (Lanza & Cooper, 2016). This method does away with the assumption that class membership is concrete or known, eliminating the bias of classification error all together (e.g. Asparouhov & Muthén, 2014). This way, the outcome variable (latent classes/trajectories) remains latent, unlike in the classify-analyze approach in which the class membership is treated as known (Vermunt, 2010). However, this method has some drawbacks. The inclusion of auxiliary variables in mixture models can be challenging because these variables may “change the nature of the heterogeneity being modeled” (page 4, Nylund-Gibson & Grimm, 2019).

In other words, once an appropriate number of classes are identified in the unconditional measurement model, the addition of covariates can shift or change the latent class structure in the structural model, so much so that it is no longer measured simply by the original latent class indicator variables but now it is also measured by the auxiliary variables (Nyland-Gibson & Mason, 2016). The shift can be so substantial that not only can parameters change from the unconditional phase, but also the number of classes. This can be especially challenging for the applied researcher as a shift in class...
structure implies that the classes are no longer based on the original latent class variable, but the covariates too, rendering the classes un-interpretable (Kamata et al., 2018; Vermunt, 2010). However, this difficulty in meaningful interpretation is less of a challenge in latent class growth analysis as the outcome is longitudinal, whereas the baseline covariates are time-invariant, unlike in latent class analysis. Nevertheless, covariates can be associated with the growth factors (intercept and slope) as well the latent class variable (Jung & Wickrama, 2008). Thus, the recommendation is to follow up the final unconditional model with a conditional model and compare results (Jung & Wickrama, 2008).

**Preliminary Data Analysis.** All analysis will be run using Mplus (Muthén & Muthén, 2019). Figures and descriptives will be derived using R. To account for the clustered nature of the data (students in schools) and attrition, the complex sampling option in Mplus (Muthén & Muthén, 2019) will be utilized with the W9C19P_2T290 weight \( n = 7,330; \) all \( N \)'s rounded to 10 as per NCES requirements) and its corresponding strata and PSU. The logit, odds \( e^{\text{logit}} \) and latent class probabilities of the unconditional and conditional models will be presented. Below is the step by step process that will be followed to carry out the analysis.

**Step 1. Exploration of data.** Bar charts describing the entire distribution of count data per grade were presented above. This confirms the high number of zeros in the data, though it is not as extreme as observed in other health science literature that looks at behaviors like smoking (e.g., Rose et al., 2006), suspensions (e.g. Morgan et al., 2019) and delinquent behavior (Reinecke, 2017). On average, an average of around 6%-11% of the data in each wave consists of zeros. To explore the functional form of the data, a
random sample of 100 was plotted from the data sample.

![Figure 1 Mean externalizing behavior from kindergarten to fifth grade. Gray line represents the predicted linear fit, blue line indicates mean behavior count for each grade.](image)

Previous studies that examined trajectories of growth for externalizing behavior have primarily fitted linear and quadratic curves to the data (e.g. Olson et al., 2017, Nagin & Tremblay, 1999), however only linear iterations have converged and been interpreted. Thus, for the sake of this prospectus, a preliminary analysis with a linear functional form will be assumed. Further inspection of the functional form will be conducted during the actual analysis to explore the best fit for the data.

**Step 2. Unconditional models.** As recommended by Jung & Wickrama (2008), first an unconditional linear growth curve model will be fitted. An unconditional model is defined as a model with only the outcome and no covariates. This will be done using both ZIP and negative binomial models to test which fits the data better. The better fitting model will be chosen to go forward with the analysis. Bayesian Information Criteria (BIC), and adjusted BIC and entropy values will be used to judge model fit and choose either a ZIP or a negative binomial distribution (Allison, 2012). Following Kreuter &
Muthén (2007) the intercept for the inflation part of the ZIP model will be fixed to zero in all classes. Additionally, the estimate of the slope is restricted to be equal across classes. For the Poisson part of the unconditional ZIP model (equation 3), without any covariates, the exponential function of the logit parameter (intercept) is simply the odds ratio for being in the particular class trajectory versus the reference class trajectory.

Once the best distribution function for the model is chosen, up to a 5-class solution will be modeled and compared to choose the most appropriate number of classes. The model with a low BIC value and a significant LMR p-value comparing the $k$ and the $k-1$ class model will be used to initially guide analysis. In other words, comparing the current model against the model with 1 less class than the current model of choice should give an LMR-RT (Lo-Mendell-Rubin likelihood ratio test) p-value less than 0.05 (Jung & Wickrama, 2008). Since, according to several authors, in mixture models a k class model is not considered nested, a $\chi^2$-difference test is not being used (Reinecke, 2017). Therefore, the BIC is being used for model comparison. Additionally, posterior probabilities for class membership will be cross-tabulated to assess the quality of class separation. Having a high probability for a single class is indicative of a clear class assignment. If average posterior class probabilities are close to one, it demonstrates a good model fit.

**Step 3 Conditional Models.** Since this study is using a one-step approach, covariates can have a significant effect on class membership, and thus unconditional models should be compared with conditional models to assess any drastic change in class structure. Similar steps as described for the unconditional model will be completed for the conditional models. The goal being to select the most parsimonious model with the
best relative fit (Jung & Wickrama, 2008; Nagin et al., 2017). A common problem when estimating complex models with several covariates is overfit (Hawkins, 2004). This happens when the complexity of your model, i.e., number of parameters being estimated and resulting degrees of freedom exceed the limits of your sample (Kline, 2015). Simply, if you apply your estimates from an overfit model to a random sample of your data, it is unlikely to fit.

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). In this study, the combination of the large data set (N= 7,330), the number of parameters being estimated, and the resulting degrees of freedom, made chances of overfit slim (Wurpts & Geiser, 2014). Nevertheless, following best practice, to assess the fit of the model and class separation, I will conduct a cross validation method known as holdout, which is preferred when using larger datasets (Gao, Calhoun & Sui, 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 will be conducted on the training set and the optimal number of classes will be obtained (Kim, 2009). Then, this model will be run on the ‘testing’ set to see if similar fit statistics and parameter estimates are returned (Kim, 2009). If there are no significant differences, this means that our model generalizes well and there is no overfit (Gao et al., 2018). The final analysis will then be run on the entire data set.

**Step 4 Trouble Shooting.** In case of the single class growth curve model fitting better than the multiple class model for behavior, these results will be written up and
discussed as final with possible reasons why the single class model was the preferred one. Another possible problem is that of non-convergence or false solutions (Reinecke, 2011). When estimating latent class growth analyses using maximum likelihood, often the algorithm may not be successful in finding the absolute maxima or solution (Boomsma, 1985). A high number of random starting values and integration points increases the likelihood that a solution, if arrived at, is valid (Boomsma, 1985). However, it also leads to an exponential increase in computation time. As the complexity of the model increases, the likelihood of convergence problems that result from software limitations also increase. In case of non-convergence, Mplus allows user defined start and integration values. Decreasing the number of random starts substantially reduces the complexity of computation and can improve chances of convergence (Muthén & Muthén, 2019). Thus, in the event of non-convergence, the default number of random start values will be adjusted to see if this solves potential problems.

In Mplus The STARTS syntax specifies the number of random sets of starting values followed by the number of final optimizations (Muthén & Muthén, 2019; Jung & Wickrama, 2008). If there are any issues relating to Mplus the author will contact the Mplus developers through the paid license obtained. Additionally, peers have agreed to serve (Andrew Jordan Thayer and Mollie Weeks) as consultant for any pressing concerns or questions. Final results will also be cross checked by making use of consultation provided to graduate students by the research methodology consultation center at the university.
Results

This section will first present steps taken for model selection, then include descriptives of the final model chosen, followed by answers to each research question. Descriptives of the weighted sample are presented in Table 2.1. Missing data percentages are presented for all covariates and the primary outcome variable at each wave. Specifically, the outcome variable had missingness that ranged from 1% to 3.69% (Table 2.1). FIML, as stated earlier, was used to handle missing data (Bhaskaran & Smith, 2014). During the analyses, I found that Mplus estimates missing data only in the dependent variables using the FIML estimation, and thus deleted cases listwise on incomplete independent variables. This led to a potential loss of 1000 cases in the main analysis. In order to avoid deleting cases listwise, MPlus authors recommended adding covariate variances into the “MODEL” command. The program then treats covariates as dependent variables in the analysis and allows consideration of all cases in the FIML estimation without changing original model specifications (Muthen, 2009). The final analytic sample therefore remained the same (N = 7,330) and no cases were dropped.

Model Selection

Model selection was first conducted on the training dataset (n = 3660) as per cross validation procedures. There were several steps taken and decisions made before a final model was selected to answer the research questions. First, I had to choose between the ZIP and negative binomial (NB) model, next the appropriate form (linear, quadratic, piecewise) had be selected. An unconditional latent curve growth model was run to compare the fit between a Zero Inflated Poisson model and Negative Binomial model. Using Allison’s (2012) recommendations to rely on the BIC (or adjusted BIC) as an
indicator of better fit, the NB model was chosen (ΔBIC 536.75). This was validated by the dispersion parameters being significant, as well as the finding that the proportion of zeros at each wave was not extremely high, ranging from 6% to 11% (Table 2.2, Figure 2.1).

Next, an appropriate form of the developmental trajectory had to be chosen. The estimated and observed mean plot (Figure 2.1) indicated that on average, behavior was stable across time, with a slight increase in teacher reported externalizing behavior in first and second grade that largely disappeared in fourth grade. Thus, apart from the slight upward curve observed in two data points, the trend was largely flat. Three unconditional NB latent curve growth models were analyzed to account for the increase in behavior in first and second grades (a) quadratic; (b) piecewise; and (c) linear. The quadratic and piecewise models did not converge, even after reducing the number of random start points. The linear unconditional model, on the other hand, converged with one error about saddle points. This error usually indicates poor model fit (Muthen, 2016), which is to be expected given that no covariates were added to this model. Based on Mplus authors’ recommendations, the saddle point warning was ignored as the model terminated normally and standard errors were returned (Muthen, 2016). Thus, overall, the linear unconditional model performed the best and this was supported by Figure 2.1. To make sure that adding covariates did not change this conclusion, models were then re-run 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, indicating improved fit.
The linear model was rerun with twice the number of random starting points to validate the replicated best log likelihood. This step makes sure that the optimal solution has been reached. Thus, to move forward with the analysis, the linear function and form were chosen. Lastly, both the slope and intercept variance parameters were significant indicating significant variance across individual externalizing behavior scores at kindergarten as well as growth. The third step was to select the appropriate number of trajectory classes to answer the research questions.

**Selecting the appropriate number of trajectory classes.** Models were run in line with the procedure described by Jung and Wickrama (2008) for a one-step approach. An unconditional, linear, negative binomial latent curve growth model was first fitted. Then, successive models were analyzed with an increasing number of classes until a 5-class solution was reached. For each unconditional model, the corresponding full model was run to make sure there were no major differences in class structure. Fit statistics for the unconditional and full models are presented in Table 2.3. The 4-class and 5-class solutions were discarded as they had poor entropy in the unconditional models, and a zero-person class in the conditional models, both indicators of poor fit. Nagin et al. (2017) recommends several fit statistics to guide the selection of the appropriate number of classes. In this study, the adjusted BIC, log likelihood, entropy, class sizes, Lo Mendel Rubin Test (LMR) and posterior class probabilities (Table 2.3) were considered. The bootstrapped LMR (BLRT), another recommended fit statistic (Kreuter & Muthén, 2008), was not calculated due to Mplus not providing this statistic for complex (i.e. weighted) sample mixture models. Corresponding plots for the two and three class solution are presented in Figure 2.3.
Ideally, a lower BIC and an increased entropy should help select the best fitting model. However, in this case the relative change in BIC was very low across the one-, two- and three-class solutions (Table 2.4). Additionally, the entropy was similar between the two- and three-class solution, indicating only slight differences in model fit. This is supported by the non-significant LRT values. The statistical evidence, therefore, suggested that all three models had similar fit statistics and were fairly equivalent. However, each model was qualitatively different and tells a different story. In the next section, theoretical and empirical evidence will be examined to choose the best fitting model (Müthen, 2003).

**Consideration of one, two, and three class solutions.** The seminal theoretical literature on the development of trajectories of externalizing problem behavior was based on Moffit’s 1993 study that examined the development of anti-social behavior in boys from early/middle childhood to adulthood. Moffit (1993) conceptualized a three-class taxonomy, (a) life course persistent problem behavior; (b) problem behavior that begins in adolescence; and (c) low to no antisocial behavior. A review of the empirical evidence following Moffit’s original paper has been largely supportive of the existence of differential trajectories of externalizing behavior and consistently points to the existence of a distinct ‘high flyers’ or persistent externalizing behavior group (Van Dulmen, Goncy,
Vest, & Flannery, 2009). Belonging to this high flyer group is related to several negative life outcomes compared to moderate and low persistent groups (Bevilacqua, Hale, Barker, & Viner, 2018). Thus, the one-class latent curve solution, which assumed all individuals in the sample came from a homogenous population, though the most parsimonious of the three models in consideration, may not be the ideal way to conceptually and empirically represent the data given the research questions.

Additionally, Figures of the 2- and 3-class solution point to distinct trajectories (Figure 2.3) that support the existence of a persistent higher problem behavior group. Given the similar BIC’s across the three solutions, and the aim of this study being to (a) investigate the association of academic achievement with membership to distinct trajectory classes, and (b) examine sociodemographic differences between these trajectory groups, the one-class solution can be safely discarded for the purpose of this study.

Choosing between the two- and three-class solution was less clear. This was because currently, there is little conceptual literature that supports two-, three- or even six-class solutions over the other when studying observed externalizing problem behavior (Walters & Rucsio, 2013), i.e., majority of literature remains empirical. Though few contest the existence of trajectories of externalizing behavior (Roisman et al., 2010), evidence on how many ‘true’ trajectories exist can vary with the way the behavior is defined (Van Dulmen et al., 2009). This is theorized to be due to Moffit’s (1993) original paper including only anti-social behavior as an outcome, and since then developmental scientists have broadened the subject of study to externalizing behavior more broadly as I have defined in this paper (Van Dulmen et al., 2009). However, what has been consistent across studies is that a lower frequency high persistent class is identified, reflecting the
relatively low rate of externalizing behavior disorders in the population (e.g. Polanczyk et al., 2015). Descriptives confirmed that this is also replicated in the current ECLS data set used for the analysis. i.e., the skewed outcome variable (Figure 2.1) of externalizing problem behavior indicated that most children do not present with externalizing problem behaviors.

Additionally, another consistent finding is that though some studies have identified two-class trajectories of externalizing behavior (e.g., Blokland & Nieuwbeert, 2005), the more commonly identified taxonomy is of high persistent, low/moderate, and no problem groups across developmental periods (e.g. Bongers et al., 2004; Reinecke, 2006; Castelao, & Kröner-Herwig, 2014). When applying the above findings to the current analysis, it is clear that the 3-class solution identified a no-problem, low-persistent and high-persistent group (Figure 2.2). The two-class solution, on the other hand, split the population into a medium and a low persistent group, which obscured the high-persistent group within the larger population, unlike the three-class model. Additionally, growth parameters of both models suggest that externalizing behavior was observed to be a stable strait across elementary school — i.e. children that enter school with high and low rates of teacher reported externalizing behavior tend to be similarly rated as they progress.

Finally, George Box’s 1987 observation that “all models are wrong, but some are useful” (Box & Draper, 1987, p. 424) is relevant to decision making when choosing between these two models. The purpose of this study, as part of the broader developmental literature, was to examine whether early academic achievement predicted membership to certain trajectory groups, for the high-persistent group in particular, in an
effort to inform early prevention and remediation (Olson, Choe & Sameroff, 2017). Though the two-class solution was able to describe the data adequately, it was unable to identify the high persistent behavior group. Whereas the three-class solution identified a high (27%) and no problem group (29%), and a majority low persistent group (43%). Thus, it aligned not only with observed and empirical data (e.g. Van Dulmen et al., 2009), but also the purpose of the study. Lastly, though the difference was small, fit statistics pointed to the conditional three-class solution having a better overall fit than the conditional two-class solution (Table 2.3).

Cross validation. Once the three-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 (Appendix C), indicating that the model building process was accurate, and supported generalization of the model. Thus, for the purpose of this study, the three-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 (Table 2.6).

Three Class Model

Means and standard deviations for each trajectory class in the three-class model are presented in Table 2.6. The high persistent (HP) intercept group mean (M = 7.39, SD = 3.50) for externalizing behavior was almost one standard deviation above the sample mean, the low persistent (LP) intercept group mean (M = 3.23, SD = 2.61) was similar to the sample mean, and the no problem (NP) group mean (M = 0.81, SD = 1.35) was
approximately one standard deviation below the sample mean \((M = 4.27, SD = 3.62)\). This trend continued through 5th grade.

The mean intercept parameters for all three trajectory classes were significant (Table 2.4). The slope intercepts were significant for the HP group \((\beta = -0.02; p < .001)\) and the LP group \((\beta = -0.01; p = .029)\) indicating a slight decrease in mean externalizing behavior from kindergarten through fifth grade for these two groups. However, the size of this effect was very small, and figures indicate that behavior was largely stable across groups in elementary school. Table 2.6 describes the results of the multinomial regression part of the latent class growth analysis, with the low persistent group as the reference group. Odds ratios are presented for ease of interpretation.

**Research Question 1**

The first research question examined whether math and reading achievement were associated with membership to differential externalizing behavior trajectory classes. i.e., did early math and reading achievement predict a student’s externalizing behavior trajectory class membership to the high persistent and no-problem group, compared to the low-persistent group? Descriptives in Table 2.4 show that the high persistent group or HP group had the lowest math and reading achievement at school entry, and the no-problem or NP group had the highest math and reading achievement at school entry. These differences were statistically significant across groups (Table 2.4); however, results from latent class growth analysis showed that after controlling for various socio-economic 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 \(\text{OR} = 1.01; 95\% \text{ CI} [1.00,1.02]; p = 0.273)\) compared to the low persistent (LP) group. Secondly,
though the $p$-value for odds ratio 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 [0.98, 0.99]; $p = .001$) and that students’ reading achievement made them almost equally likely to belong to either group. Similarly, though significant, students’ math achievement was weakly associated with classification to the HP (OR = 1.02; 95% CI [1.01, 1.03]; $p = 0.002$) or NP group (OR = 1.01; 95% CI [1.00, 1.02]; $p = .138$) compared to the LP group. The odds ratios being close to one discouraged meaningful interpretation of the significant $p$-value.

Overall, neither children’s math nor reading achievement scores at school entry were associated with class membership to the high persistent and no problem behavior groups compared to the low persistent group, after controlling for relevant covariates.

Nevertheless, class membership was significantly related to several other variables included in the model.

**Research Question 2**

The second research question examined the child and family 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 high persistent trajectory group, and almost half as likely (OR = 0.48; 95% CI [0.42, 0.55]; $p < .001$) than girls to belong to NP group compared to the low persistent group, after controlling for covariates. 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 than white students (OR = 0.41; 95% CI [0.28, 0.60]; $p < .001$) to belong to the NP group 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 compared to the LP group compared to White students.

Socioeconomic status was also related to trajectory membership. Since latent class growth analysis is essentially a multinomial regression and socioeconomic status was treated as continuous, every one-unit increase in the socioeconomic scale does not result in a linear increase in odds. Instead, the increase in odds follows a log function. Therefore, to allow for easier interpretation, a brief statement of the relationship between socioeconomic class of a child and trajectory class will be described, followed by a comparison of the odds ratio of a child belonging to the lowest versus the highest socioeconomic class. Overall, students who belonged to families who were of higher socioeconomic status were less likely to be categorized in the HP group, compared to the LP group (OR = 0.60; 95% CI [0.51, 0.69]; \( p < .001 \)). For example, if a student belonged to the highest level of socioeconomic status compared to lowest socioeconomic status, their odds of belonging in the HP group instead of the LP group decreased by more than half (\( \Delta \) OR = 0.58). Socioeconomic status was also significantly related to belonging to the NP group compared to the LP group, indicating that children belonging to higher socioeconomic status was 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 higher proportion of minority students was not associated with class membership after controlling for covariates (Table 2.5). This result was replicated
in the between-class descriptives (Table 2.6), where students in each class belonged to schools that had similar mean percentages of minority students.

**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 much lower odds of belonging to the high persistent group compared to the low persistent group. For example, if a teacher changed the rating of a student from an average of 1 (never displays behavior) on the scale to 4 (always displays behavior), this would decrease the odds of belonging to the high persistent group compared to the low persistent group by 0.38. Similarly, if a teacher changed the rating of a child from an average of 1 (always displays behavior) on the scale to 4 (never displays behavior), this would increase the odds of a child belonging to the NP group compared to the LP group by 0.36.

Findings were similar for the approaches to learning variable. An increase in approaches to learning teacher ratings was associated with a decreased likelihood of belonging to the HP group 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 likelihood of a student belonging to the NP group compared to the LP group (OR = 1.75; 95% CI [2.50, 2.04]; \( p < .001 \)). For example, if a teacher changed the rating of a child from an average of one (never displays behavior) to four (always displays behavior), this would decrease the odds of belonging to the high persistent group, compared to the low persistent group, by almost half (\( \Delta \text{OR} = 0.47 \)). Similarly, if a teacher changed the rating of a child from an average of one (never displays behavior) to four (always displays behavior), this would increase the odds of a student belonging to the NP group
compared to the LP group by more than seven ($\Delta$ OR=$7.53$) times. The working memory measure and language measure were not related to class membership after controlling for covariates. Similarly, the presence of an IEP in kindergarten was also not related to class membership (Table 2.5).

**Discussion**

This study, using a person-centered latent class growth approach, studied 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 were examined that were associated with trajectory membership. Identifying whether the relationship is stronger for certain subgroups of students is critical to inform prevention efforts, especially in kindergarten and first grade, as schools could target intervention efforts on the mechanism to improve both academic and behavioral functioning (Masten & Cicchetti, 2010). Overall, three class trajectories of externalizing behavior were identified; a high persistent group (HP), a low persistent group (LP) and a no problem group (NP). A majority of students belonged to the LP and NP group, indicating that most students did not have teacher-reported behavioral difficulties in elementary school. Importantly, neither early math nor reading achievement was associated with trajectory membership. However, being a boy, identifying as Black, belonging to lower socioeconomic status, lower teacher reported attention, and lower teacher reported approaches to learning skills were all associated with classification in the HP compared to the LP group. Overall, students in the HP group had the lowest academic achievement, teacher rated attention, and approaches to learning scores in kindergarten.
Early Academic Achievement and Externalizing Behavior

Overall, results showed that neither math nor reading achievement were related in a meaningful way to class membership. This was found across the high persistent and no problem group when compared to the low persistent group, suggesting that the relationship was similar at different levels of externalizing behavior. There were significant differences for means in reading and math student achievement across groups, however, the latent class growth analysis showed that after controlling for relevant covariates, academic achievement did not predict class membership. This finding implies that the variation in mean academic achievement at school entry observed across behavior trajectory groups was not associated with the level of externalizing behavior observed in students. Although students in the HP group had worse reading and math scores compared to the LP and NP groups, this difference was likely due to variables other than teacher-reported externalizing behavior, i.e. the effect of academic achievement on trajectory membership was likely attenuated by the presence of other covariates. Though there are developmental (e.g., Sameroff, 2009) as well as behavioral theories (Mcintosh, 2005) that directly link low academic achievement to the development of externalizing behavior, results from this study do not support any predictive relations between the two domains. Findings from Study 1 also support this result, as the few studies included in the review that examined the relationship in the direction of low achievement to externalizing behavior found mostly non-significant effects. For example, socio-demographic factors (Burt & Roisman, 2010) and executive functioning (Morgan et al., 2018) predicted externalizing behavior rather than achievement. In other studies, inattention (e.g., Rabiner, 2000) fully mediated the
relationship, and cognitive ability (e.g., Masten, 2005), when accounted for, removed any significant associations between the two domains.

Findings also lend support to Hinshaw’s (1992) antecedent model theory (Figure 1.1), in which the observed comorbidity between lower academic achievement and externalizing behavior is explained by theoretical antecedent variables. Specifically, malleable variables like teacher-reported inattention and approaches to learning seem to be promising variables to examine in their relation to the development of externalizing behavior and achievement. Future studies should directly test these pathways with appropriate time ordered variables to further support this theory.

**Inattention and Approaches to Learning**

The results from this study were similar to those that report no association between academic achievement and externalizing behavior when specific covariates are accounted for, specifically, attention (e.g. Rabiner et al., 2000; Metcalfe et al., 2013). Students whose teachers reported having a higher number of inattentive behaviors were more likely to belong in the high persistent group than the low persistent group, and those with higher number of attentive behaviors were also more likely to be in the no-problem group than the low persistent group. A large body of literature supports this result, as several authors have suggested that low attentional control is associated with increased frequency of externalizing behavior (e.g., Morris et al., 2014), and that within the domain of externalizing behavior, oppositional-defiant behavior has the strongest overlap with inattention (e.g., Lahey, 2008). Given the evidence of inattention’s associations with both academic achievement and externalizing behavior in Study 1, coupled with this study’s results, it supports evidence that it could play an antecedent or mediating role in the
relationship between the two domains (Hinshaw, 1992; Hoffman, 2018). Additionally, given that inattention also strongly predicts academic achievement when externalizing behavior is controlled for (e.g., Horwood and Fergusson, 1995; Gray, Dueck, Rogers & Tannock, 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 (Grimm et al., 2010). It is important to note, however, there is some contradictory literature suggesting only hyperactive symptomology is associated with externalizing behavior, and not inattentive symptoms (e.g. Ahmed & Hinshaw, 2017).

Similar to inattention, low teacher ratings in approaches to learning was associated with membership to the high persistent group, compared to the low persistent group, and higher frequencies of approaches to learning was associated with children belonging to the no problem group compared to the low persistent group. The measure used in this study was developed as a dimension of school readiness (Kagan, Moore, & Bradekamp, 1995) that encompasses behavior critical to classroom success like transitioning, task persistence, and following rules (Razza, Martin, & Brooks-Gunn, 2015, p. 757). Results aligned with previous studies demonstrating that these skills are not only highly predictive of academic achievement (e.g. Bulotsky-Shearer et al., 2011), but externalizing behavior as well (Razza et al., 2015). It is possible that students who have higher externalizing behavior in kindergarten also display a combination of other problem behaviors like inattention and poor approaches to learning (Grimm et al., 2010), and this contributes to the consistent ratings of externalizing behaviors through elementary school and even adulthood (Reef et al., 2010).
Similarly, the numbers reversed score, a measure of working memory, showed discrepancies in the mean performance amongst groups, but was not a significant predictor of class membership. It is likely that other covariates, known to be related to working memory and externalizing behavior like inattention (Rogers et al., 2011; Shoemaker et al., 2013), contributed to this result. This hypothesis is supported by the small to medium correlations working memory shared with both inattention and the externalizing behavior outcomes in this study (Table 2.3). Thus, the primary malleable predictors of class membership were not direct measures of working memory or academic achievement, but rather teacher-observed behavioral measures, specifically inattention and approaches to learning.

**Language and internalizing behavior.** Recent longitudinal meta analytic results examining the relationships between language and behavior are of small but negative associations (e.g., Chow & Wehby, 2018). The findings from this study did not support this trend, but a potential reason could be that most students performed very well on the language measure used in ECLS (Table 2.1); thus, resulting in a lack of variation and power to predict membership to the externalizing behavior groups, especially after covariates added in the model (Cohen, 2013). Second, researchers have theorized that externalizing behavior behaves as functional communication in response to impaired language ability (Hollo & Chow, 2015). However, since language is only one way in which communication occurs, using problem behavior as successful communication may be an infrequent event (Chow, Ekholm & Coleman, 2018). Lastly, teacher-report of externalizing behavior is subjective, and direct standardized measures of classroom
behavior could have stronger associations with oral language than teacher recall of observed behavior (Chow et al., 2018).

Internalizing behavior problems have been related to the development of externalizing behaviors in several theoretical and empirical studies (e.g., Masten et al., 2005). Most theory is contextualized around developmental cascades (Masten & Cicchetti, 2010), in which early externalizing behavior is believed to play a role in the development of internalizing behavior problems like depression and anxiety as children age due to the multiple failures in competence domains, including behavior (Sameroff, 2009). In this study, however, students whose teachers rated them has having higher internalizing behavior were more likely to belong to the no-problem group, compared to the low-persistent group. This is in line with empirical studies that showing early internalizing behavior, is usually associated with adolescent externalizing behavior, but not externalizing behavior in childhood (e.g. Bornstein et al., 2010; Weeks et al., 2017).

Socio-Demographic Characteristics

There were several student and family characteristics associated with trajectory membership. Specifically, a student’s sex and race were the two strongest predictors to trajectory membership even after controlling for cognitive and behavioral measures. Boys, compared to girls, were almost twice as likely to belong to the high persistent trajectory group and almost half as likely to belong to the no problem group compared to the low persistent group. The finding that male children display more observable externalizing behavior has been established several times (e.g. Else-qvest et al., 2006; Hyde, 2014) and has both genetic and environmental causes (e.g., Björkqvist, 2018). Overall, it is generally agreed upon that even though levels of aggression can be similar
across sexes, boys, on average, display more direct and physical behaviors, whereas girls tend to display more indirect aggressive behaviors (Hyde, 2014). These findings are in line with the current study’s results, as teachers are more likely to observe externalizing behaviors that were disruptive and aggressive (De los Reyes et al. 2015).

**Race.** Student race was also a predictor of trajectory membership. Even after controlling for covariates, Black students’ behavior ratings by teachers made them almost twice as likely as their white peers to belong to the HP class trajectory and half as likely to belong to the NP trajectory compared to the LP trajectory class. This trend was not observed for Hispanic children or the other race variable (Table 2.6). Though the low persistent trajectory class was proportionate in terms of race, the no-problem trajectory class again had a disproportionally small number of Black children, with only 11% of Black children belonging to this group, compared to Hispanic (32.52%), White (31.72%), and other race (34.35%) children. These results, combined with the multinomial regression, suggest that teachers were consistently rating a large proportion of Black students as having higher externalizing problem behaviors, compared to other races, and specifically compared to white children. This difference is not unsubstantial 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 (King et al., 2004; Bronfenbrenner, 1994). At the student-teacher 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). This extends to white teachers’ expectations of Black students’ academic achievement (e.g. Redding, 2019). This discrepancy can be conceptualized within critical race theory, and teachers’ ratings as a reflection of teachers’ and larger societies’ pathologizing of Black students, and boys especially, behavior (Watts & Erevelles, 2004). Further, teacher ratings cannot be seen divorced from school context. At the school-level, Black students are more likely to attend schools with fewer resources in lower socioeconomic neighborhoods (Owens et al., 2016), and as a result are more likely to be exposed to inappropriate discipline practices (Smolkowski et al., 2016), higher student-teacher conflict (Collins et al., 2016), and neighborhood violence (Vélez et al., 2016); all of which are associated with maladaptive socioemotional responses, including higher externalizing behavior (Skalicka et al., 2016; Teske et al., 2011). These schools also tend to have teachers that are poorly trained and highly stressed (Simon & Johnson, 2015), resulting in conditions that exacerbate implicit bias and activation of harmful stereotypes (Langer, 2009). Moreover, scholars argue that the systemic sociopolitical violence committed on the Black community by white supremacist institutions and processes has led to generational trauma, lack of opportunity, and poverty (Smedley & Smedley, 2015). Given that results indicated lower socioeconomic status also predicted membership to the high problem stable group, it compounds the risk of Black children being exposed to environmental risk factors that are associated with poorer mental health (e.g. Smedley, Smith & Nelson, 2003). Further, color blind policies and discipline frameworks in schools erase the context of culture, implicit bias, and natural responses to trauma (Artiles et al., 2010; Annamma et al., 2017).
Accordingly, we must temper our conclusions about the disproportionate representation of Black students in the high persistent trajectory within the context of the socio-political history of the United States (Sullivan, 2017). Specifically, teachers’ ratings of Black student’s behavior must be seen through the lens of the historical and political context within which these behaviors are occurring and how they are being perceived. At the same time, to truly prevent these inequitable outcomes, educators and policy makers could focus on what structural inequities are resulting in these outcomes rather than exclusively focusing on the individual child’s or teacher’s behavior (Artiles et al., 2010). In conclusion, we cannot shy away from acknowledging that results indicate that systemic racism, both at the individual-, school-, and institutional-level could affect Black students’s behavior, teachers’ perceptions of their behavior, and how schools, as institutions, respond to behavior.

Implications for Practice

Several implications can be drawn from the results of this study. First, results do not support a unidirectional relationship between early academic achievement and externalizing behavior. Though unidirectional and bi-directional developmental theories of the causal relations between academic achievement and externalizing behavior exist (e.g., Patterson et al., 1990; Miles & Stipek, 2006), the results of this study suggest that a student’s early academic achievement does not make them more or less likely to have 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 that are antecedent causes to both externalizing behavior and academic achievement (Algozzine et al., 2011; Hinshaw, 1992).
Second, results imply that interventions 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 (Lane, Oakes and Menzies, 2014). In other words, the prevention of behavioral problems should depend on evidence-based instruction or intervention in behavior rather than expecting gains in one domain to carry over to the other (Corcoran, et al., 2018; Sugai & Horner, 2009). This is especially important for schools that have fewer resources as they are more likely to serve children that are exposed to trauma (Abramovitz & Mingus, 2014) and have less experienced teachers (Ronfeldt et al., 2013). Thus, these schools will likely need a more trauma-informed framework to improve behavioral outcomes instead of purely achievement focused curricula (Chafoleus et al., 2016).

Third, results support evidence that teacher-reported externalizing behavior is largely stable through elementary school (e.g., Olson et al., 2017). This underscores the importance of early intervention to prevent behaviors from developing before kindergarten (e.g., Dodge et al., 2010). Interventions targeted at improving early school readiness skills has been shown to prevent development of later externalizing behavior problems (Schindler et al., 2015), 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 (Blair & Raver, 2015), could result in greater reduction of problem behavior (Duncan et al., 2018).
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’s behavior start early and remain largely stable, supporting other findings that have found biased observations of behavior can begin as early as preschool (e.g., Gillam, 2016). Early externalizing behavior, as perceived by teachers, has long term negative consequences, including higher probability of being identified for special education services (e.g., Kulkarni & Sullivan, 2019), risk of drop out, and even adult psychopathology (Reef et al., 2011). These outcomes are often exacerbated by increased likelihood of disciplinary consequences, of which Black children are disproportionally affected by (Bottiani et al., 2017). Disciplinary consequences strongly predict both juvenile and adult incarceration and are said to play a role in strengthening the school to prison pipeline (Skiba et al., 2014), which again disproportionally affects Black children, and especially Black boys (Annamma et al., 2014).

As evidence suggests that Black teachers rate Black students’ behavior more favorably compared to their white counterparts (Gates & Blick 2013) and also have higher expectations of their Black students (Fox, 2016), schools could support the intentional recruitment of Black educators to help counterbalance the negative bias that non-Hispanic white teachers might have when rating Black students’ behavior (e.g., Redding, 2019). Further, implicit bias training may not be enough to change behavior (Gillam et al., 2016). Rather, researchers suggest implementing cultural competency training for teachers to provide context to student’s behaviors and prevent misinterpretation (Blake et al., 2016), potentially interrupting the school to prison
pipeline by reducing discipline referrals (Allan & White-Smith, 2014). Lastly, 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).

In conclusion, the study results suggest that teacher-reported externalizing behavior is fairly stable through elementary school, and to improve behavioral outcomes of all children, early intervention targeting attention and school readiness behaviors like approaches to learning, rather than academic achievement, could be more effective than just externalizing behavior alone (Grimes et al., 2010). Additionally, early academic achievement may not have the cross-domain protective effects as previously theorized (e.g., Masten, 2005), and thus schools should not expect improvement of behavior or prevention of the development of externalizing behavior by focusing solely on academic achievement (Algozzine et al., 2011).

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 warrant disaggregated data. Thus, results may not generalize to students who belong to these racial and ethnic categories. Second, the measures of behavior included in this study are short form, and though they have high reliability for the sample, consisted of fewer items compared to gold standard long form measures like the Achenbach’s System of Empirically Based Assessment (ASEBA; Achenbach & Rescorla, 2001). Even though the specific brief assessments of behavior
used in this study (SSRS; Gresham & Elliot, 1990) are considered to have adequate validity evidence to be used to measure teacher ratings externalizing behavior (e.g., Gresham et al., 2010), the clinical utility of the ratings are potentially reduced (Achenbach et al., 2016). Additionally, 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., 2019), thus future research should include middle and high school students from appropriate datasets.

Further, the latent class growth analysis in this study was conducted via the Mplus (Muthén & Muthén, 2018) computer program, and thus the analysis was constrained within its limitations. The BLRT (Feng & McCulloch, 1996) test was not available for use for complex samples which limited the ability to use it when comparing the one-, two-, and three-class models. Additionally, as a one step, rather than a three step, approach (Vermunt, 2010) was used, users depend on the program to select the comparison class, which is usually assigned to the class with the highest number of cases. In this study, it was the low persistent group, and thus the regression was carried out with this group as the comparison group. As a result, estimates and interpretation could change if the no-problem group was selected as the comparison group instead.

Finally, through the statistical model selection process, the three-class trajectory model was chosen based on application of theory and previous empirical evidence; however, the one-class model also sufficiently described the data. This, and the variation of trajectory classes found in other studies (e.g. Van Dulmen et al., 2009) underscores that latent classes or trajectories depend on theory, are largely empirical, and always contain classification error (Muthén, 2003). In this case, the results showed that
unobserved heterogeneity existed in the data but was satisfactorily modeled by the latent curve growth model. Thus, the three-class model was only one way to describe the data. Even though the entropy and posterior classification probabilities reported in this study were high (Nyland-Gibson & Choi, 2018), no student perfectly belonged to any one trajectory, and thus inferences made from the results of this study must be contextualized within the constraints of the sample and the statistical method itself (Vermunt, 2010).

Chapter 4 Discussion

The purpose of this dissertation was to investigate the relationship between academic achievement and externalizing behavior, in an effort to reveal any causal or predictive relations between the two domains. Low competency in either domain is associated with several negative long-term consequences, with evidence suggesting that failure in both can lead to cumulatively worse outcomes (Broidy et al., 2003). Empirical and theoretical literature on the relationship between behavior and achievement is largely ensconced within the framework of a developmental cascade (Masten & Cicchetti, 2010; Sameroff, 2009) — i.e. failure in one domain can trigger outcomes that result in increased probability of failure in the related domain. Though there have been several studies that have examined this relationship, there has been little consensus on the validity of the relationship (Algozzine et al., 2011). It is important to note that without manipulation of variables it is impossible to truly confirm any potential causal relations between these variables (Holland, 1986), and the nature of variables themselves makes it near impossible to design a study that does (Kazdin, 2007). To examine causal relations between these two variables, therefore, it is recommended to rely on summaries of best evidence (Shadish, 2010), with the caveat that true causal inference can only be inferred
if a variable, when manipulated or changed, shows a corresponding change in the outcome (Kraemer, 2001). The last systematic review on the topic was completed in 1992 by Stephen Hinshaw, who examined the evidence for causal relations between the two domains. He concluded that methodological flaws in study designs and inappropriate covariates precluded any robust conclusions about a causal relationship (Hinshaw, 1992).

However, since then, and despite limited evidence, several interventions in schools have been designed based on the conceptualization that improving performance in either domain prevents the development of problems in the other (e.g., Bradshaw et al., 2009). The purpose of this dissertation, therefore, was two-fold: to systematically review the literature on the evidence of causal relations between externalizing behavior and academic achievement, and to design a study that examined whether early academic achievement was associated with the development of externalizing behavior, using recommendations from the findings of the first study. Results from this dissertation should inform efficient resource allocation, intervention design, and implementation of future interventions in schools.

The results of the first study did not support the inference of causal relations between the academic achievement and externalizing problem behavior. Indeed, results suggested that there was not a reliable predictive association either. In other words, neither reliable unidirectional nor bidirectional associations exist between externalizing behavior and academic achievement. Results lent credence to Hinshaw’s (1992) antecedent model (Figure 1.1), where observed comorbidity found between the two domains are better explained by common antecedent causes (e.g., inattention, language, and cognitive ability; Study 1). When examining associations, all studies in the review
assumed that the sample was homogenous, and that there was no unobserved heterogeneity in the way externalizing behavior develops across children. This assumption is contrary to popular developmental theory and empirical evidence that identifies at least three distinct trajectories of externalizing behavior development and distinct risk and protective factors (e.g. Olson et al., 2017).

Thus, the second study in this dissertation chose to fill these gaps in the literature by designing a study that could test the unidirectional pathway between early academic achievement and externalizing behavior, with appropriate covariates and a person-centered design that did not assume all children were drawn from the same population. Interestingly, results from study two were similar to results in study one. There was no convincing evidence that early academic achievement was directly related to the development of externalizing behavior in elementary aged school children (e.g., Rabiner, 2000). Rather, other antecedent variables like inattention and approaches to learning were the strongest malleable factors that predicted trajectory membership. Finally, socio-demographic factors like a child’s sex, race, and socioeconomic status — rather than an indicator of school context — emerged as robust predictors to trajectory membership.

**Implications for Practice**

The results of this dissertation suggest overwhelmingly that the comorbidity of low achievement and high externalizing behavior that is often observed in children is due to underlying antecedent causes (e.g. Hinshaw, 1992; Algozzine et al., 2011). This conclusion has several implications for schools and educators, specifically in terms early intervention and prevention. It is crucial for intervention developers and educational leaders to have an accurate representation of the targeted developmental domains and potential outcomes; in this case, whether intervening on teacher reported disruptive
behavior can indeed result in improved academic performance (Bradshaw, et al., 2009). Firstly, the findings from Study 1 imply that schools should have realistic expectations of universal social emotional learning programs (SEL) that are advertised as interventions targeting behavioral outcomes, but also as leading to gains in academic achievement (CASEL, 2013). Policy and practice (e.g. Every Student Succeeds Act, 2015) have largely encouraged implementation of SEL as an intervention to improve academic achievement without a strong research base to justify the investment of monetary and human resources (Algozzine et al., 2011). This is supported by a recent meta-analysis of randomized clinical trials that showed SEL programs significantly improved emotional and behavioral outcomes, but academic achievement was not substantially affected (Corcoran et al., 2018). Implementation of SEL programs expecting gains in achievement could result in a potential loss of instructional time better spent on educational experiences or interventions with a greater likelihood of producing positive outcomes for affected students.

Secondly, results from Study 2 reiterated that students’ early academic achievement does not predict classification to high or low externalizing behavior groups, and that focusing on behaviors like inattention (e.g. del Mal Bernard et al., 2017) and approaches to learning (Razza et al., 2015) could lead to larger gains in preventing the development of externalizing behavior. Investing in interventions that target the problem behavior, rather than rely on cross domain effects, could be the most efficient use of resources for both universal and more intensive programs (Algozzine et al., 2011). The slight decrease, but largely stable teacher ratings of externalizing behavior through elementary school has been a consistent finding in several other studies (e.g. Olson et al.,
2017), indicating that children who enter kindergarten with higher teacher-rated externalizing behaviors continue to be rated similarly through elementary school.

Viewed from an ecological framework, this finding implies that early childhood is a critical period of vulnerability for the development of externalizing behaviors (Olson et al., 2020), and results underscore the importance of preschool interventions to prevent problem behavior before it becomes intractable (Dodge et al., 2010). Results also suggest that, in addition to individual-level behavior contributing to the relatively stable teacher ratings through elementary school, school and classroom contexts associated with externalizing behavior problems (e.g., poor student-teacher relations, Collins et al., 2016; and exclusionary discipline practices, Smolkowski et al., 2016), are also potentially remaining unchanged as children progress through elementary school. The complex interplay of these variables, including the negative bias of teachers against students of color, especially Black children (Halberstadt et al., 2020), could contribute to the stable teacher perceptions of behavior.

Thus, schools and policy could not only focus on student-level interventions targeting the behavior, but also consider the contexts in which these behaviors occur (Artiles et al., 2010; Bates & Glick, 2013). For example, teacher-level interventions that promote cultural competency training (Blake et al., 2016), as well as systems-level interventions (e.g., providing disadvantaged families wraparound support including childcare) could improve behavioral outcomes for young children (Dodge et al., 2017). A final example of a systems-level intervention could be Promise Neighborhoods that offer communities health, employment, school resources, and childcare in an effort to improve
both academic and behavioral outcomes of children (Gordon, Jean-Louis, & Obiora et al., 2017).

**Future Directions**

Findings from this dissertation contribute to existing evidence that the relationship between achievement and externalizing behavior is not as strong as previously believed (Algozzine et al., 2011). Rather, other related variables like inattention could completely account for any comorbidity between these two domains (e.g. Rabiner, 2000). This finding is similar across different levels of externalizing behavior. Previous studies within developmental psychology have focused on internalizing and externalizing behavior in relation to academic achievement (e.g. Burt & Roisman, 2010); however, future studies could use similar methodological designs to investigate the role of inattention in developmental cascades and the development of problem behavior as well as low academic achievement across childhood and adolescence. Similarly, domains of school readiness that are closely related to both inattention and approaches to learning like self-regulation (Blair & Raver, 2015) could be investigated to elucidate their role in the development of externalizing behavior and inform early intervention (Duncan et al., 2018).

Findings also support the implementation of a blended model of service that targets both behavior and academics for the best student outcomes (Lane et al., 2014). There has also been some research that suggests adopting more trauma-informed frameworks in response to maladaptive behavior (Chafouleas, Johnson, & Overstreet, 2016) could lead to fewer disciplinary referrals (Blitz, Anderson, & Saastamoinen, 2016) as well as improved teacher ratings of academic and behavioral competence (Mendelson et al., 2015). That said, more research is needed to validate the efficacy of these claims.
and identify which school practices lead to improved outcomes (Maynard et al., 2019).

Lastly, in addition to investigating student and teacher factors associated with positive academic and behavioral outcomes, future studies should investigate the contribution of systems to student development and potential student outcome disparities (Ungar & Leibenberg, 2013).

A final area for future research could be the validity of teacher ratings of externalizing behavior when assessing racial and ethnic minority children. Findings from this study and other reviews (e.g., Redding, 2019) suggest that white teachers’ biases against Black students’ behavior and academic potential are pervasive and negatively affect their ratings of these students’ behaviors. These findings could have broader implications for teacher-student relationships when there is a race mismatch (Gershenson, Holt, & Papageorge, 2016), and considerations of how teacher bias contributes to disproportionality in discipline referrals (Smolkowski et al., 2016), as well as special education eligibility and placement decisions (Waitoller, Artiles, & Cheney, 2010).

Additionally, future studies that examine inter-rater reliability between parent and teacher report of externalizing behavior could study whether bias or racism influence agreement scores (e.g., Kang & Harvey, 2019) in addition to contextual differences between home and school as suggested previously (De Los Rayes et al., 2015).

**Conclusion**

This dissertation was dedicated to the examination of the relationship between externalizing behavior and academic achievement in children. The purpose was to study in detail any potential predictive and causal relations in order to inform early intervention and prevention. The first study systematically reviewed the literature in order to reveal potential causal relations, if any, whereas the second study aimed to fill specific
methodological and empirical gaps found in reviewed literature. Overall, results from Studies 1 and 2 suggest that any comorbidity found between these two variables, likely has other causes, lending support to Hinshaw’s (1992) antecedent model. Some malleable variables that could potentially underlie both academic achievement and externalizing behavior are attention (e.g., Rabiner et al., 2000), approaches to learning (e.g., Razza et al., 2015) and language development (Chow & Wehby, 2018). However, more research, appropriate data, and methodological designs are needed to generalize these findings and investigate possible causal relations.

More importantly, findings have specific implications for schools, and suggest that academic achievement may not have the preventative cross domain effect on externalizing behavior as many believe, and vice versa (Algozzine et al., 2011). Thus, focusing prevention and intervention efforts on both domains is likely to lead to the best outcomes for students (Lane et al., 2014). Results from Study 2, specifically, add to evidence (e.g. Gates & Blick, 2013) that suggests Black students are rated by their teachers as having consistently higher rates of externalizing behavior compared to their white peers. This finding indicates that in addition to individual level intervention to reduce externalizing behaviors, teacher- and school-level interventions to reduce bias in behavior ratings could be beneficial (e.g., Blake et al., 2016). Lastly, the findings in this dissertation underscore that the development of externalizing behavior and academic competence in children is a complex, dynamic process (Masten & Cichetti, 2010) and that a child’s academic competence and potential to succeed academically, can likely be viewed independent of their perceived ‘problem’ behavior.
References


Allen, Q., & White-Smith, K. A. (2014). “Just as bad as prisons”: The challenge of dismantling the school-to-prison pipeline through teacher and community


http://static.squarespace.com/static/513f79f9e4b05ce7b70e9673/t/526220de4b00a92c90436ba/1382687245993/2013-casel-guide.pdf


https://doi.org/10.1080/00220671.2015.1039113

https://doi.org/10.1016/j.edurev.2017.12.001

https://doi.org/10.1016/j.jsp.2012.09.005

why language is central for learning development. *Educational Researcher, 39*,

and behavioral antecedents to early adolescent involvement with antisocial peers.

https://doi.org/10.1111/j.1469-7610.2008.01985.x

early childhood programs and policies on educational outcomes in elementary

Bornstein, & T. Leventhal (Eds.), *Handbook of child psychology and
developmental science, volume four; ecological settings and processes* (7th ed., pp.

leaves behind: The roles of IQ and self-control in predicting standardized


school achievement and delinquent behavior. *Journal of Abnormal Child
Psychology, 23*(2), 183-199.

equation modeling. In G.R Hancock & R.O. Mueller (Eds.), *Structural Equation
Modeling: A second course* (269-314). Greenwich, CT: Information Age
Publishing.

on teacher expectations and recommendations. *AERA Open, 2*(1),
2332858415623758. https://doi.org/10.1177/2332858415623758

and implications of rapid reviews. *Implementation Science, 5*, 1-10.
https://doi.org/10.1186/1748-5908-5-56

classification to treatment outcome prediction. *CNS Neuroscience &
Therapeutics, 24*, 1037-1052. ttps://doi.org/10.1111/cns.13048

parenting and child adjustment in young children. *Journal of Clinical Child and
Adolescent Psychology, 33*, 347-358.
https://doi.org/10.1207/s15374424jccp3302_16

of student–teacher demographic match on teacher expectations. *Economics of


https://doi.org/10.1017/S0954579410000337


https://doi.org/10.1177/1063426608324724


from systematic reviews of risk factors: The Cambridge Quality Checklists.


http://www.statmodel.com/discussion/messages/22/4448.html


http://www.statmodel.com/discussion/messages/12/23080.html?1549502418


https://doi.org/10.1111/1467-8624.00086


https://doi.org/10.1080/21683603.2018.1425169


https://doi.org/10.1007/s10802-007-9208-2


https://doi.org/10.1080/13803395.2014.921284


https://doi.org/10.1017/s0954579410000076


https://doi.org/10.1080/09297049.2010.544648


https://doi.org/10.1080/10543400600719384


https://doi.org/10.1002/sim.2331


https://doi.org/10.1002/9781119125556.devpsy113


https://doi.org/10.1093/aje/kwi042


https://doi.org/10.1037/e609692011-001


https://doi.org/0009-3920/2006/7701-0006


https://doi.org/10.1177/0143034312472761


https://doi.org/10.1093/acprof:oso/9780195310313.003.0014


https://doi.org/10.1017/S0954579410000283


https://doi.org/10.1093/pan/mpq025


Table 1.1

*Descriptives of Studies Included in Systematic Review*

<table>
<thead>
<tr>
<th>Characteristics of sample</th>
<th>Range</th>
<th>Did not Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Female</td>
<td>20%-63%</td>
<td>0</td>
</tr>
<tr>
<td>Percent White</td>
<td>11%-95%</td>
<td>3</td>
</tr>
<tr>
<td>Percent Black</td>
<td>9.5% to 50%</td>
<td>5</td>
</tr>
<tr>
<td>Percent Hispanic</td>
<td>3% to 24.3%</td>
<td>8</td>
</tr>
<tr>
<td>Percent Asian/PI</td>
<td>1%-40%</td>
<td>12</td>
</tr>
<tr>
<td>Percent did not complete high school*</td>
<td>5.2%-11%</td>
<td>9</td>
</tr>
<tr>
<td>Percent defined low income**</td>
<td>0%-33.3%</td>
<td>6</td>
</tr>
<tr>
<td>Average age at initial data</td>
<td>6.2 years</td>
<td>3</td>
</tr>
<tr>
<td>collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average age at follow up</td>
<td>12.8 years</td>
<td>2</td>
</tr>
<tr>
<td>Sample Size</td>
<td>35-8950</td>
<td>0</td>
</tr>
<tr>
<td>Year of data collection</td>
<td>1985-2013</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note:* *parents of children in the sample. **defined in studies as percent on free reduced lunch, annual income <20,000, below 25th percentile in a national sample*
<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample Characteristics (sample size)</th>
<th>Child age at each wave</th>
<th>Externalizing Variables defined at T1</th>
<th>Academic Variables defined at T1</th>
<th>Externalizing Variables defined at T2</th>
<th>Academic Variables defined at T2</th>
<th>Findings</th>
<th>Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burt &amp; Roisman (2010)</td>
<td>Stratified sampling from 10 research sites across the NICHD SECCYD ($n = 1,160$)</td>
<td>24 months, 54 months, 1st, 3rd, 5th grade and 15 years</td>
<td>Maternal report of delinquent and aggressive behaviors-CBCL</td>
<td>WJ R</td>
<td>Maternal report of delinquent and aggressive behaviors-CBCL</td>
<td>All subtests of the WJ-R</td>
<td>1st grade AA→3rd grade Ext ($\beta=-1.0$)</td>
<td>IQ, sex, income needs ratio, maternal sensitivity</td>
</tr>
<tr>
<td>Hurry et al. (2018)</td>
<td>Children with reading difficulties from 63 schools in east England ($n = 258$)</td>
<td>6 yrs, 9 months into school year, 10 yrs</td>
<td>Teacher report of conduct disorders, emotional disorders and hyperactivity-CBQ</td>
<td>BAS, Neale Analysis of Reading, OSELA, The Oddities Test</td>
<td>Teacher report of conduct problems, emotional symptoms, and hyperactivity-inattention-SDQ</td>
<td>NFER-Nelson Group Reading Test 6-12, The Parallel Spelling Test</td>
<td>Ext 6yrs → AA 10yrs (conduct: $\beta=-.14$, p=.034, hyperactivity: $\beta=-.01$, p=.49), AA→E ($\beta=0.04$, $\beta=0.01$, p&gt;.05)</td>
<td>Age, sex, English as an Additional Language, IQ</td>
</tr>
<tr>
<td>Metcalfe et al. (2013)</td>
<td>Children with externalizing problems and controls who participated in a study of early development of ADHD and ODD ($n = 199+60$)</td>
<td>3yrs, 6 yrs</td>
<td>BASC PRS Hyperactivity, Aggression, Attention Problems subscales</td>
<td>KSEALS Composite, Vocabulary, Numbers, Letters, and Words tests</td>
<td>BASC PRS Hyperactivity, Aggression, Attention Problems subscales</td>
<td>WIAT–II–A; word reading, math calculation, and spelling.</td>
<td>No effect of aggression or hyperactivity with academic achievement, $\beta=-0.10$, $\beta=-0.10$, p&gt;.05</td>
<td>Sex, race, inattention, SES, family stress, maternal depression</td>
</tr>
</tbody>
</table>
### Morgan et al. (2017)
Nationally representative dataset - ECLS-K-2011 \((n=8,920)\)
- K, 2nd grade
- Teacher report of Externalizing subscale of the SSRS
- Standardized IRT test of reading and math
- Teacher report of externalizing subscale of the SSRS
- Standardized test of reading and math
- \(K \text{ Ext} \rightarrow 2\text{nd grade reading and math AA} (\beta=0.03, p<.05)\)

### Rabiner et al. (2000)
Sites included Durham, Nashville, Seattle, and rural, central Pennsylvania. 4 schools chosen from “high crime” areas. \((n=396)\)
- K, 1st, 2nd grade.
- Overactivity scale from the Child Attention Problems Scale - Teacher.
- Externalizing scale from the TRF
- Letter-Word Identification, Sight word identification -WJ.
- Hyperactive-impulsive factor of the ADHD Rating Scale
- Letter-Word Identification, Sight word identification, Passage Comprehension -WJ.
- No significant effect of externalizing/hyperactivity on reading achievement and vice versa.

### Weeks et al. (2016)
Children followed bi-annually as part of the National Longitudinal Study of Children and Youth in Canada \((n=6425)\)
- 4 yrs, 6 yrs, 8 yrs, 10 yrs, 12 yrs, 14 yrs, 16 yrs
- Items derived from the externalizing scale of OCHS-R and CBCL – parent and self-report
- CAT/2 for Math
- Items derived from the externalizing scale of OCHS-R and CBCL – parent and self-report
- CAT/2 for Math
- AA 8yrs→ Ext 10yrs, AA at 10yrs→ Ext 12yrs, \(\beta =0.08, p<.05\), Ext 8yrs→ AA 10yrs, Ext 12yrs→ AA 14yrs, \(\beta =0.07, p<.05\).

---

**Note:** NICHD SECCYD = National Institute of Child Health and Human Development Study of Early Child Care and Youth Development, ADHD = Attention Deficit Hyperactivity Disorder, BAS = British Ability Scales, BASC = Behavior Assessment Scale for Children, CBCL = Child Behavior Checklist, CBQ = Child Behaviour Questionnaire, CD = Conduct Disorder, DSM IV = Diagnostic Statistical Manual, 4th edition, GCSE = General Certificates of Secondary Education, LD= Learning Disability, LA= Low Achievement, ODD= Oppositional Defiant Disorder, OSELA = The Observation Survey of Early Literacy Achievement, PACS = the Parental Account of Childhood Symptoms, PACAS = the Parental Account of Child and Adolescent Symptoms, PIAT= Peabody Individual Achievement Test, RD=Reading Disability, SDQ = Strengths and Difficulties Questionnaire, SSRS = Social Scale Rating System, TRF = Teacher Report Form, WJ III = Woodcock Johnson, 3rd edition, WRAT = Wide Range Achievement Test, SAT = Scholastic Aptitude Test. AA=Academic Achievement.
# Table 1.3

**Studies that Examined Unidirectional Models**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample frame (sample size)</th>
<th>Child age at initial and follow up</th>
<th>Externalizing Variables measured at T1</th>
<th>Achievement Variables measured at T1</th>
<th>Externalizing Variables measured at T2</th>
<th>Achievement Variables measured at T2</th>
<th>Findings</th>
<th>Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brennan et al. (2012)</td>
<td>Low SES subsample recruited from the Women, Infants, and Children Nutrition Programs (n = 566)</td>
<td>2yrs, 7.5 years</td>
<td>Parent report inattention, hyperactivity-impulsivity, aggression and oppositionality - ECBI</td>
<td>NA</td>
<td>Parent report - hyperactivity-impulsivity, aggression - CBCL</td>
<td>Letter-Word Identification, Math Calculation, and Spelling of the WJ III</td>
<td>Agg at 4-5 years → AA at 7.5 years, β = -0.16*, intervention not related to increase in AA, β = -0.06</td>
<td>General language quotient, inattention, sex, race, parent education</td>
</tr>
<tr>
<td>Breslau et al. (2009)</td>
<td>Community samples of low and normal birth weight children from two hospitals in MI (n = 713)</td>
<td>6 years, 17 years</td>
<td>Teacher report of delinquent behavior; aggressive behavior - TRF</td>
<td>NA</td>
<td>Teacher report of inattention delinquent behavior, aggressive behavior - TRF</td>
<td>WJ-R- Basic Reading, Reading Comprehension, Calculation and Applied Problems</td>
<td>No effect of delinquent behavior on later AA, β = -0.05, -0.01, p &gt; 0.05</td>
<td>IQ, inattention, birthweight, sex, mother’s education</td>
</tr>
<tr>
<td>Chadwick et al. (1999)</td>
<td>6- and 7-year-old children attending primary school (n = 2462)</td>
<td>6 - 9 yrs, 16 - 18 yrs</td>
<td>Rutter parent and teacher child behavior questionnaire, PACS</td>
<td>Neale Analysis of Reading Ability, Grading Word Spelling Test</td>
<td>PACAS</td>
<td>GCSE</td>
<td>No significant effect of hyperactivity on persistence of reading disability Early</td>
<td>Perinatal adversity, neurologic problems</td>
</tr>
<tr>
<td>Duncan et al. (1999)</td>
<td>NICHD</td>
<td>4.5 years, 15</td>
<td>Maternal report</td>
<td>W-J Reading-</td>
<td>NA</td>
<td></td>
<td>IQ,</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Sample Characteristics</td>
<td>Age</td>
<td>Measures</td>
<td>Findings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------------------------</td>
<td>-----</td>
<td>----------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duncan et al. (2007) *</td>
<td>SECCYD (n = 1160)</td>
<td>125 years</td>
<td>delinquent and aggressive behaviors-CBCL</td>
<td>Externalizing → AA at 5th grade (ß=.05, p) inattention, social skills, internalizing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duncan et al. (2007) *</td>
<td>Children of women from the NLTY study. Stratified sampling, Black and Hispanic women oversampled. (n = 1756)</td>
<td>5-6 years, 13-14 years</td>
<td>Parent report of “Head strong” behavior, antisocial-ABPC PIAT Reading Recognition</td>
<td>PIAT- Reading and Math K ext reading at age 13 (ß=−.05) Language, internalizing, inattention</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitzpatrick &amp; Pagani (2013)*</td>
<td>Children randomly selected from a stratified sampling of infants born in Quebec. (n = 960)</td>
<td>K, 4th grade</td>
<td>Parent reported hyperactive behavior, behavioral problems, physical aggression-SBQ Number Knowledge Teacher reported hyperactive behavior, behavioral problems, physical aggression-SBQ</td>
<td>CAT/2 in math No effect of hyperactive behavior on overall AA at 4th grade (unstandardized ß = .02, p&gt;.05) Classroom engagement, receptive vocabulary, number knowledge, behavioral problems (internal, external and prosocial), sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gray, et al. (2014)</td>
<td>At risk sub sample from a larger sample of age and gender stratified random sampling of infants (n = 359)</td>
<td>25 months, 8 years</td>
<td>Maternal report of inattention, overactivity, aggression-ITSEA</td>
<td>NA WJ III Broad reading composite No effect of aggression (ß₀ =−.05, ß₁=−.16 p&gt;.05), hyperactivity (ß₀ =.03, ß₁=0.26 p&gt;.05) on 2nd grade reading.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Survey Type</td>
<td>Initial Range</td>
<td>Measure of Externalizing Behavior</td>
<td>Academic Achievement Measures</td>
<td>Findings</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------</td>
<td>--------------------------------------</td>
<td>---------------</td>
<td>-----------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kremer et al. (2016)</td>
<td>PSID - longitudinal survey</td>
<td>Initial range of 3-6 yrs, 18 yrs</td>
<td>Behavior Problem Index (BPI) to measure the incidence and severity of child externalizing behavior problems</td>
<td>NA</td>
<td>Early Ext→ later AA, Internalizing, age, race, sex, developmental delay, mother’s education, SES, birthweight</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massetti et al. (2008)</td>
<td>ADHD probands and controls recruited from psychiatric clinic and school samples</td>
<td>PreK, K, 1st, 2nd, 3rd, 4th, 6th, 7th grade</td>
<td>Parent report - DISC-2.3 and Teacher Report-DBD Rating Scale</td>
<td>WJ - Letter-Word Identification test, Applied Problems Test, the Passage Comprehension test</td>
<td>Hyperactive symptom group was not different from control group for reading, math, β = 3.26, p = .16, β = 0.40, z = 0.18, p = .36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IQ, age, inattention, income, internalizing disorders.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salla et al. (2016)</td>
<td>QLSCD - longitudinal survey of singleton newborns representative of those born in Quebec (Canada) between 1997 and</td>
<td>1.5-5 yrs, 6 yrs, 7 yrs, 8 yrs, 9 yrs, 12 yrs</td>
<td>Parent report of hyperactivity and inattention</td>
<td>NA</td>
<td>Results on province-wide ministerial exams for reading, writing, and mathematics were not significantly associated with AA.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gender, anxiety/depressive symptoms, opposition, aggression</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaughn &amp; Haager (1994)</td>
<td>Subsample of larger community sample of two urban schools (n = 29)</td>
<td>K, 1st, 2nd, 3rd, 4th, 5th grade</td>
<td>Teacher report of CD, socialized aggression, attention problems on RBPC</td>
<td>Three groups: LD, LA and HA.</td>
<td>Teacher report of CD, socialized aggression, attention problems on RBPC</td>
<td>SAT achievement scores</td>
<td>No significant difference in behavior problems across groups from 1st to 3rd grade.</td>
<td>None reported</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------------------------------------</td>
<td>-----------------------------</td>
<td>---------------------------------------------------------------</td>
<td>------------------------</td>
<td>---------------------------------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Bussing et al. (2012)</td>
<td>Random Sampling from a Florida Public School data base. (n = 112+87+23)</td>
<td>3rd to 10th grade</td>
<td>Parent and Teacher Report; three groups: ADHD, subclinical ADHD- 4 or 5/9 symptoms, meets DSM IV criteria and control Maternal report of conduct problems and attention deficit behaviors based on Rutter and Connors questionnaires</td>
<td>NA</td>
<td>NA</td>
<td>Florida School District state test, based on the SAT, GPA, retention, Graduation rate.</td>
<td>No relation of ADHD status on math, reading, $\beta = -5.65$ $p=.40$, $\beta =11.25$, $p=.06$ IEP status, gender, FRL status, Hollingshead Four factor index,</td>
<td></td>
</tr>
<tr>
<td>Fergusson &amp; Horwood (1995)</td>
<td>Subsample of birth cohort from Christchurch Health and Development Study who were resident in the Christchurch urban region</td>
<td>8 yrs, 13 yrs, 15 yrs</td>
<td>Self-report, parent report of delinquent behaviors- SRED, police contacts</td>
<td>TOSCA</td>
<td>AA at 13yrs not related to delinquency at 15 years, r=0.005 IQ, inattention, early conduct problems</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Initial Data collection in elementary school*
<table>
<thead>
<tr>
<th>Study</th>
<th>Sample Description</th>
<th>Age at Data Collection</th>
<th>Data Collection Methods</th>
<th>Parameter(s) Linking Ext to AA</th>
<th>IQ, Gender, SES, Age, Inattention</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fergusson et al. (1993)</td>
<td>Birth cohort of children from the Christchurch Health and Development Study (n = 761)</td>
<td>6 yrs, 8 yrs, 10 yrs, 13 yrs</td>
<td>Maternal and teacher reports of conduct/oppositional and attention deficit/hyperactive behaviors based on Rutter and Connors questionnaires</td>
<td>Parameter(s) linking Ext to AA was small and non-significant (β = - .02, p &gt; .80).</td>
<td>IQ, gender</td>
</tr>
<tr>
<td>Masten et al. (2005)</td>
<td>Sample from urban Minneapolis neighborhood (n = 205)</td>
<td>8 yrs-12 years, 15-18 yrs</td>
<td>Parent report – aggression and delinquency on the CBCL*</td>
<td>NA</td>
<td>IQ, sex, parenting quality, SES</td>
</tr>
<tr>
<td>Miller et al. (2014)</td>
<td>Students from 19 public schools in a southeastern urban district. Students who had the lowest reading scores were selected (n = 110)</td>
<td>1st and 3rd grade</td>
<td>Teacher report on the SWAN for attention and hyperactive/impulsive behavior</td>
<td>Sound matching, phonemic decoding, listening comprehension, sight word-TOWRE</td>
<td>NA</td>
</tr>
<tr>
<td>Rapport, et al. (1999)</td>
<td>Community sample from Hawaii (n = Mean age of 10.67 years to 13.67</td>
<td>Parent report inattention symptomology,</td>
<td>NA</td>
<td>SAT scores in reading and Math</td>
<td>No relationship between IQ, SES, age, inattention</td>
</tr>
<tr>
<td>Study</td>
<td>Sample Description</td>
<td>Age</td>
<td>CD Symptomology</td>
<td>Raters</td>
<td>Math and Reading Scores</td>
</tr>
<tr>
<td>-------</td>
<td>-------------------</td>
<td>-----</td>
<td>----------------</td>
<td>--------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Realmuto, et al. (2000)</td>
<td>Community sample of 22 suburban elementary schools screened for disruptive behavior and controls (n = 204)</td>
<td>Grades 2-5 to grades 5-8</td>
<td>Teacher, parent scores on the Conners index and BASC, peer scores on RCP</td>
<td>NA</td>
<td>RCP and BASC scores</td>
</tr>
<tr>
<td>Rennie, et al. (2014)</td>
<td>ADHD symptomology and controls from a larger sample of children from a southern California public school district and private school district (n = 17+34)</td>
<td>1st-3rd grade, 2nd to 4th grade, and 3rd to 5th grade. High and Low ADHD symptomology groups based on two teacher reports on the Conners index.</td>
<td>Reading, comprehension and Arithmetic subtests (WRAT-III)</td>
<td>NA</td>
<td>Reading, comprehension and Arithmetic subtests (WRAT-III)</td>
</tr>
</tbody>
</table>


Groups based on ADHD DSM-IV criteria, Reading and spelling composite of the PIAT, teacher report on the DBRS and DICA IV.

Compared to Controls:
- RD group had increase in hyperactivity symptoms, Δd=0.2. However, no difference in hyperactivity between groups. (δ=0.2).
- RD group had higher special education needs. Behavioral difficulties rated by teacher and parent.
- RD group had higher rates of conduct issues.
Note: * studies that controlled for initial measures of their outcomes in their models. Masten et al. (2005) collected measures at both time points, however their academic latent measure at follow up did not include a standardized measure of achievement and thus was excluded from the review for consistency of reporting. Willcutt et al (2007) did not present time ordered analysis with standardized academic outcomes, however the t tests testing the AA to hyperactivity relation could be interpreted in a time ordered way. ADHD = Attention Deficit Hyperactivity Disorder, BAS = British Ability Scales, BASC = Behavior Assessment Scale for Children, CBCL = Child Behavior Checklist, CAT/2 = Canadian Achievement Test, CD = Conduct Disorder, DSM IV = Diagnostic Statistical Manual, 4th edition, NKT = the Number Knowledge Test, ODD = Oppositional Defiant Disorder, PIAT = Peabody Individual Achievement Test, PPVT = Peabody Picture Vocabulary Test, RD = Reading Disability, SAT = Scholastic Aptitude Test, SBQ = Social Behavior Questionnaire, SDQ = Strengths and Difficulties Questionnaire, SRED = Self-report Early Delinquency Scale, SSRS = Social Scale Rating System, SWRD = specific word reading difficulty, TOSCA = Test of Scholastic Abilities, TRF = Teacher Report Form, WJ III = Woodcock Johnson, 3rd edition, WRAT = Wide Range Achievement Test, WRT = Woodcock Reading Mastery Test-Revised. If effect or statistical estimates not present, then these were not present in the data made available by authors.

Table 1.4

<table>
<thead>
<tr>
<th>Authors</th>
<th>QI %</th>
<th>Statistical Design</th>
<th>Group Variable</th>
<th>Covariates found significant</th>
<th>Path</th>
<th>Effect Magnitude (+/-/ne)</th>
<th>Developmental Span</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burt &amp; Roisman (2010)</td>
<td>88.88</td>
<td>SEM</td>
<td>No</td>
<td>IQ, early AA, Early E, maternal sensitivity, SES and social competence</td>
<td>E→AA</td>
<td>ne</td>
<td>Preschool and elementary school</td>
</tr>
<tr>
<td>Massetti et al. (2008)</td>
<td>83.33</td>
<td>Regression Behavior</td>
<td></td>
<td>IQ, internalizing, inattentive type</td>
<td>E→AA</td>
<td>ne</td>
<td>Elementary and middle school</td>
</tr>
<tr>
<td>Duncan</td>
<td>80.55</td>
<td>Regression No</td>
<td></td>
<td>IQ, language,</td>
<td>E→AA</td>
<td>β = -0.03, +0.05</td>
<td>Preschool and Elementary,</td>
</tr>
<tr>
<td>Study</td>
<td>Sample Size</td>
<td>Method</td>
<td>Outcomes</td>
<td>Model</td>
<td>Parameters</td>
<td>Findings</td>
<td>School Level</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-------------</td>
<td>--------</td>
<td>---------------------------------------</td>
<td>-------</td>
<td>------------</td>
<td>----------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Brennan et al. (2012)</td>
<td>77.77</td>
<td>SEM</td>
<td>inattention, early AA</td>
<td>No</td>
<td>E→AA</td>
<td>β = -0.23</td>
<td>Preschool and elementary school</td>
</tr>
<tr>
<td>Breslau et al. (2009)</td>
<td>77.77</td>
<td>Regression</td>
<td>IQ, inattention, maternal education</td>
<td>No</td>
<td>E→AA</td>
<td>ne</td>
<td>Elementary and High School</td>
</tr>
<tr>
<td>Masten et al. (2005)</td>
<td>77.77</td>
<td>SEM</td>
<td>IQ, Early E Vocab, Early AA, WM, IC, race, gender, IEP Early AA, SES</td>
<td>No</td>
<td>AA→E</td>
<td>ne</td>
<td>Elementary and high school</td>
</tr>
<tr>
<td>Morgan et al. (2018)</td>
<td>75</td>
<td>Regression</td>
<td>IQ, inattention, parent education E→AA</td>
<td>No</td>
<td>β = -0.03</td>
<td></td>
<td>Elementary school</td>
</tr>
<tr>
<td>Metcalfe et al. (2013)</td>
<td>72.22</td>
<td>SEM</td>
<td>Early E, SES Early AA, SES</td>
<td>No</td>
<td>AA→E</td>
<td>ne</td>
<td>Pre school</td>
</tr>
<tr>
<td>Rabiner et al. (2000)</td>
<td>69.44</td>
<td>Regression</td>
<td>IQ, inattention, parent involvement</td>
<td>No</td>
<td>E→AA</td>
<td>ne</td>
<td>Elementary School</td>
</tr>
<tr>
<td>Realmuto et al. (2000)</td>
<td>66.67</td>
<td>Regression</td>
<td>IQ, age</td>
<td>No</td>
<td>E→AA</td>
<td>R² = 0.30</td>
<td>Elementary and middle school</td>
</tr>
<tr>
<td>Bussing et al. (2012)</td>
<td>61.11</td>
<td>Regression</td>
<td>Special education status</td>
<td>No</td>
<td>E→AA</td>
<td>ne</td>
<td>Elementary, middle and high school</td>
</tr>
<tr>
<td>Rennie et al. (2014)</td>
<td>61.11</td>
<td>ANOVA</td>
<td>IQ, age, WM, RI</td>
<td>E→AA</td>
<td>β = 0.45,0.51</td>
<td></td>
<td>Elementary and middle school</td>
</tr>
<tr>
<td>Kremer et al. (2016)</td>
<td>61.11</td>
<td>Regression</td>
<td>Age, Sex, Race</td>
<td>E→AA</td>
<td>β = 0.02</td>
<td></td>
<td>Preschool/ kindergarten and high school</td>
</tr>
<tr>
<td>Rapport et al. (1999)</td>
<td>55.55</td>
<td>SEM</td>
<td>IQ, WM, classroom performance</td>
<td>No</td>
<td>E→AA</td>
<td>ne</td>
<td>Elementary and high school</td>
</tr>
</tbody>
</table>
Gray et al. (2014) 55.55 Latent Growth Curve Analysis No Age, inattention E→ AA ne Preschool and elementary school

Willcutt et al. (2007) 38.89 Regression Behavior, Reading Unclear E→ AA ne Middle and high school

Vaughn & Haager (1999) 33.33 ANOVA Reading scores Unclear AA→E ne Middle and high school

Note: ANOVA = Analysis of Variance, SEM = Structural Equation Modeling, E= Externalizing behavior problems, Behavior, AA=Academic Achievement, IQ= Intelligence Quotient, WM= Working Memory, RI= Response Inhibition, SES= Socioeconomic status, ne= No effect

Table 1.5
Quality Indicator Scores in Brief

<table>
<thead>
<tr>
<th>Study</th>
<th>Description of Sampling (/6)</th>
<th>Description of Measurement (/4)</th>
<th>Statistical Methodology, Research Design (/6)</th>
<th>External Validity and Limitations (/2)</th>
<th>Score (/18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burt &amp; Roisman (2010)</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Massetti et al. (2008)</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>15</td>
</tr>
<tr>
<td>Duncan et al. (2007)</td>
<td>5</td>
<td>2.5</td>
<td>5</td>
<td>2</td>
<td>14.5</td>
</tr>
<tr>
<td>Breslau et al. (2009)</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Brennan et al. (2012)</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Masten et al. (2005)</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Morgan, et al. (2018)</td>
<td>5</td>
<td>3.5</td>
<td>5</td>
<td>1</td>
<td>13.5</td>
</tr>
<tr>
<td>Metcalfe et al. (2013)</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Rabiner et al. (2000)</td>
<td>2</td>
<td>4</td>
<td>5.5</td>
<td>1</td>
<td>12.5</td>
</tr>
<tr>
<td>Realmuto et al. (2000)</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>12</td>
</tr>
</tbody>
</table>
Bussing et al. (2012) | 4 | 1 | 3 | 2 | 11
Rennie et al. (2010) | 4 | 3 | 3 | 1 | 11
Kremer et al. (2016) | 4 | 1 | 4 | 1 | 11
Rapport et al. (1999) | 4 | 1 | 4 | 1 | 10
Gray et al. (2014) | 3 | 2 | 3 | 2 | 10
Willcutt et al. (2007) | 2 | 2 | 2 | 1 | 7
Vaughn & Haager (1994) | 2 | 1 | 2 | 1 | 6

*Note:* Quality Indicators derived from Tooth, Ware and Banes (2010) checklist for cohort studies and Hinshaw’s (1992) requirements for enabling causal inferencing.
### Table 2.1.

**Descriptives of Weighted Sample**

<table>
<thead>
<tr>
<th></th>
<th>Mean/Pct (SD)</th>
<th>Range</th>
<th>SE</th>
<th>% missing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Child Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>51.43</td>
<td>0.007</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>51.71</td>
<td>0.023</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>13.36</td>
<td>0.167</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>24.82</td>
<td>0.020</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>10.10</td>
<td>0.011</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Birth weight in pounds</td>
<td>6.85 (1.31)</td>
<td>5-8</td>
<td>0.021</td>
<td>10.89</td>
</tr>
<tr>
<td>Reading IRT K Fall</td>
<td>53.38 (11.31)</td>
<td>0-205</td>
<td>1.50</td>
<td>0.65</td>
</tr>
<tr>
<td>Math IRT K Fall</td>
<td>35.46 (11.44)</td>
<td>0-206</td>
<td>1.53</td>
<td>0.26</td>
</tr>
<tr>
<td>Numbers reversed score K Fall</td>
<td>98 (16.82)</td>
<td>45-200</td>
<td>3.32</td>
<td>0.40</td>
</tr>
<tr>
<td>Attention K spring</td>
<td>3.03 (0.88)</td>
<td>1-5</td>
<td>0.001</td>
<td>3.90</td>
</tr>
<tr>
<td>Approaches to Learning K spring</td>
<td>2.92 (0.67)</td>
<td>1-4</td>
<td>0.005</td>
<td>3.03</td>
</tr>
<tr>
<td>Internalizing symptoms K spring</td>
<td>1.50 (0.48)</td>
<td>1-4</td>
<td>0.002</td>
<td>1.02</td>
</tr>
<tr>
<td>Total language score K fall</td>
<td>18.73 (2.75)</td>
<td>0-20</td>
<td>0.11</td>
<td>1.02</td>
</tr>
<tr>
<td>IEP in Kindergarten spring</td>
<td>8.8</td>
<td>0.000</td>
<td>8.84</td>
<td></td>
</tr>
<tr>
<td>Percent of minority students in school</td>
<td>14.64 (22.35)</td>
<td>0-100</td>
<td>6.41</td>
<td>0.22</td>
</tr>
<tr>
<td>Externalizing in K</td>
<td>3.69 (3.78)</td>
<td>0-18</td>
<td>0.17</td>
<td>1.05</td>
</tr>
<tr>
<td>Externalizing 1st grade</td>
<td>4.27 (3.62)</td>
<td>0-18</td>
<td>0.16</td>
<td>3.69</td>
</tr>
<tr>
<td>Externalizing 2nd grade</td>
<td>4.27 (3.70)</td>
<td>0-18</td>
<td>0.16</td>
<td>0.68</td>
</tr>
<tr>
<td>Externalizing 3rd grade</td>
<td>4.09 (3.65)</td>
<td>0-18</td>
<td>0.16</td>
<td>0.60</td>
</tr>
<tr>
<td>Externalizing 4th grade</td>
<td>3.88 (3.57)</td>
<td>0-18</td>
<td>0.15</td>
<td>0.98</td>
</tr>
<tr>
<td>Externalizing 5th grade</td>
<td>3.82 (3.58)</td>
<td>0-18</td>
<td>0.15</td>
<td>1.05</td>
</tr>
<tr>
<td><strong>Family Characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Socio-economic status</td>
<td>-0.09 (0.77)</td>
<td>-2.33, 2.60</td>
<td>0.007</td>
<td>0.37</td>
</tr>
<tr>
<td>Parent Education Level</td>
<td>4.56 (1.82)</td>
<td>1-8</td>
<td>0.040</td>
<td>0.29</td>
</tr>
</tbody>
</table>
Note. Weighted estimates reported. Unweighted N=7330. All unweighted counts rounded to 10 as per NCES regulations.
Table 2.2

Proportion of zeros in each wave of outcome variable.

<table>
<thead>
<tr>
<th>Wave</th>
<th>Proportion of zeros</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kindergarten</td>
<td>11.8</td>
</tr>
<tr>
<td>First</td>
<td>5.3</td>
</tr>
<tr>
<td>Second</td>
<td>6.0</td>
</tr>
<tr>
<td>Third</td>
<td>6.5</td>
</tr>
<tr>
<td>Fourth</td>
<td>7.1</td>
</tr>
<tr>
<td>Fifth</td>
<td>6.8</td>
</tr>
</tbody>
</table>
Table 2.3.

Correlation Matrix of Variables used in Analysis

|                  | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    | 13    | 14    | 15    | 16    | 17    | 18    | 19    | 20    | 21    | 22    | 23    |
|------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| **Male**         |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |
| White            | 0.01  | 0     | -0.01 | 0.09  | -0.04 | 0.03  | -0.5  | -0.22 | -0.24 | 0.03  | -0.01 | 0.11  | 0     | 0.19  | 0.21  | 0.22  | 0.26  | 0.25  | 0.26  | 0     | 0     |       |       |
| Black            | 0.00  | -0.41 | -0.59 | -0.35 | 0.1   | 0.16  | 0.25  | 0.18  | 0.06  | 0.08  | 0.32  | 0.02  | -0.3  | -0.06 | -0.03 | -0.03 | -0.04 | -0.06 | -0.06 | 0.3   | 0.33  |       |       |
| Hispanic         | 0.00  | -0.59 | -0.23 | 1     | -0.19 | 0.01  | -0.18 | -0.22 | -0.15 | -0.02 | -0.03 | 0     | -0.39 | -0.01 | -0.12 | -0.03 | -0.03 | -0.06 | -0.04 | -0.03 | -0.06 | -0.34 | -0.32 |       |
| Other race       | -0.01 | -0.35 | -0.13 | -0.19 | 1     | -0.05 | 0.106 | 0.07  | 0.05  | 0     | 0.00  | 0.01  | -0.03 | -0.01 | -0.03 | -0.08 | -0.03 | -0.04 | -0.04 | -0.05 | -0.04 | 0.05  | 0.08  |       |
| Birth weight     | -0.01 | -0.1  | -0.11 | 0.01  | -0.05 | 1     | 0.05  | 0.11  | 0.1   | 0.07  | 0.08  | 0.06  | 0.03  | -0.04 | -0.1  | -0.06 | -0.03 | -0.03 | -0.02 | -0.02 | -0.04 | -0.02 | 0.05  |       |
| Reading IRT K Fall | -0.04 | 0.16  | -0.06 | -0.18 | 0.06  | 0.05  | 1     | 0.73  | 0.39  | 0.28  | 0.31  | -0.09 | 0.32  | -0.12 | -0.05 | -0.1  | -0.09 | -0.09 | -0.09 | -0.09 | -0.07 | 0.35  | 0.39  |       |
| Math IRT K Fall  | 0.03  | 0.25  | -0.14 | -0.22 | 0.07  | 0.11  | 0.73  | 1     | 0.49  | 0.33  | 0.36  | -0.14 | 0.35  | -0.16 | -0.14 | -0.13 | -0.12 | -0.11 | -0.11 | -0.11 | 0.4   | 0.4   | 0.43  |       |
| Numbers reversed K Fall | -0.05 | 0.18  | -0.12 | -0.15 | -0.05 | 0.1   | 0.39  | 0.49  | 1     | 0.27  | 0.29  | -0.13 | 0.25  | -0.22 | -0.1  | -0.14 | -0.11 | -0.12 | -0.12 | -0.12 | 0.26  | 0.29  |       |       |
| Attention K Fall | -0.22 | 0.06  | -0.07 | -0.02 | 0     | 0.07  | 0.28  | 0.33  | 0.27  | 1     | 0.88  | -0.24 | 0.1   | -0.17 | -0.02 | -0.56 | -0.42 | -0.41 | -0.38 | -0.36 | -0.31 | 0.09  | 0.13  |       |
| Approaches K Spring | -0.24 | -0.08 | -0.1  | -0.03 | 0.01  | 0.08  | 0.31  | 0.36  | 0.29  | 0.88  | 1     | -0.33 | 0.13  | -0.19 | -0.05 | -0.6  | -0.42 | -0.41 | -0.39 | -0.37 | -0.34 | 0.12  | 0.15  |       |
| Internalizing symptoms K Spring | 0.03  | 0     | 0.03  | 0     | -0.03 | -0.06 | -0.09 | -0.14 | -0.13 | -0.24 | -0.33 | 1     | -0.04 | 0.09  | 0.01  | 0.28  | 0.11  | 0.11  | 0.09  | 0.09  | 0.07  | -0.05 | -0.08 |       |
| Language K Fall  | -0.01 | 0.32  | 0.05  | -0.39 | -0.01 | 0.03  | 0.32  | 0.35  | 0.25  | 0.1   | 0.13  | -0.04 | 1     | -0.09 | 0     | -0.02 | 0.01  | 0.02  | 0.03  | 0.02  | 0.03  | 0.31  | 0.29  |       |
| IEP K Spring     | 0.11  | 0.02  | 0.01  | -0.01 | -0.03 | -0.04 | -0.12 | -0.16 | -0.22 | -0.17 | -0.19 | -0.09 | -0.09 | 1     | 0     | 0.1   | 0.11  | 0.09  | 0.08  | 0.07  | 0.06  | -0.07 | -0.08 |       |
| % minority students | 0     | -0.3  | 0.66  | -0.12 | -0.08 | -0.1  | -0.05 | -0.14 | -0.1  | -0.02 | -0.05 | 0.01  | 0     | 0     | 0     | 0.09  | 0.07  | 0.11  | 0.08  | 0.11  | 0.13  | -0.11 | -0.19 |       |
| Ext K             | 0.19  | -0.06 | 0.15  | -0.03 | -0.03 | -0.06 | -0.1  | -0.13 | -0.14 | -0.56 | -0.6  | 0.28  | -0.02 | 0.1   | 0.09  | 0.59  | 0.56  | 0.53  | 0.5   | 0.45  | -0.04 | -0.1  |       |
| Ext 1st grade    | 0.21  | -0.03 | 0.12  | -0.03 | -0.04 | -0.03 | -0.09 | -0.11 | -0.11 | -0.42 | -0.42 | 0.11  | 0.01  | 0.11  | 0.07  | 0.59  | 1     | 0.61  | 0.58  | 0.55  | 0.49  | -0.05 | -0.1  |       |
| Ext 2nd grade    | 0.22  | -0.03 | 0.16  | -0.06 | -0.04 | -0.03 | -0.09 | -0.12 | -0.12 | -0.41 | -0.41 | 0.11  | 0.02  | 0.09  | 0.11  | 0.56  | 0.61  | 1     | 0.61  | 0.58  | 0.54  | -0.05 | -0.1  |       |
| Ext 3rd grade    | 0.26  | -0.04 | 0.15  | -0.04 | -0.04 | -0.02 | -0.09 | -0.11 | -0.11 | -0.38 | -0.39 | 0.09  | 0.03  | 0.08  | 0.08  | 0.53  | 0.58  | 0.61  | 1     | 0.60  | 0.55  | -0.06 | -0.1  |       |
### Externalizing Behaviors and Academic Achievement

|       | 0.25  | -0.06 | 0.17  | -0.03 | -0.05 | -0.02 | -0.09 | -0.11 | -0.12 | -0.36 | -0.37 | -0.09 | 0.02  | 0.07  | 0.11  | 0.50  | 0.55  | 0.58  | 0.60  | 1  | 0.6  | -0.07 | -0.12 |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|---|------|-------|-------|
| Ext 4th grade |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |   |     |       |       |
| Ext 5th grade | 0.26  | -0.06 | 0.2   | -0.06 | -0.04 | -0.04 | -0.07 | -0.1  | 0.12  | -0.31 | -0.34 | 0.07  | 0.03  | 0.06  | 0.13  | 0.45  | 0.49  | 0.54  | 0.55  | 0.60  | 1  | -0.06 | -0.12 |
| Socio-economic status | 0     | 0.3   | -0.06 | -0.34 | 0.05  | 0.02  | 0.35  | 0.40  | 0.26  | 0.09  | 0.12  | -0.05 | 0.31  | -0.07 | -0.11 | -0.04 | -0.05 | -0.05 | -0.06 | -0.07 | -0.06 | 1  | 0.82  |       |
| Parent Education | 0     | 0.33  | -0.14 | -0.32 | 0.08  | 0.05  | 0.39  | 0.43  | 0.29  | 0.13  | 0.15  | -0.08 | 0.29  | -0.08 | -0.19 | -0.1  | -0.1  | -0.1  | -0.1  | -0.12 | -0.12 | 0.82 | 1     |
| Mean | 51.43 | 51.71 | 13.36 | 24.82 | 10.10 | 6.85  | 53.38 | 35.46 | 98    | 3.03  | 2.92  | 1.50  | 18.73 | 8.8   | 14.64 | 3.69  | 4.27  | 4.27  | 4.09  | 3.88  | 3.82  | -0.09 | 4.56 |
| SD  | 1.31  | 11.31 | 11.44 | 16.82 | 0.88  | 0.67  | 0.48  | 2.75  |       |       |       |       |       |       |       |       |       |       |       |       | 3.65  | 3.57  | 3.58  | 0.77  | 1.82 |

**Note.** Weighted correlations presented. Unweighted N=7330. Ext=Externalizing, K=Kindergarten.
Table 2.4

*Fit statistics for one, two, and three class models*

<table>
<thead>
<tr>
<th></th>
<th>-LI</th>
<th>Adj.BIC</th>
<th>%ΔBIC</th>
<th>LRT</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Conditional Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 class</td>
<td>290997.49</td>
<td>583225.073</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>2 class</td>
<td>293703.71</td>
<td>588540.256</td>
<td>0.90</td>
<td>P=0.74</td>
<td>0.85</td>
</tr>
<tr>
<td>3 class</td>
<td>291555.36</td>
<td>584357.993</td>
<td>-0.72</td>
<td>P=0.77</td>
<td>0.82</td>
</tr>
<tr>
<td><strong>Unconditional Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 class</td>
<td>98851.02</td>
<td>197764.967</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>2 class</td>
<td>101008.23</td>
<td>202079.398</td>
<td>2.12</td>
<td>P=0.74</td>
<td>0.84</td>
</tr>
<tr>
<td>3 class</td>
<td>99160.14</td>
<td>198400.369</td>
<td>-1.85</td>
<td>P=0.77</td>
<td>0.81</td>
</tr>
</tbody>
</table>

*Note.* Fit statistics for 4 and 5 class models not reported will be available in supplementary material.
Table 2.5

*Mean parameter estimates and classification probabilities*

<table>
<thead>
<tr>
<th></th>
<th>Unweighted N</th>
<th>Intercept (SE)</th>
<th>Slope (SE)</th>
<th>CP Class 1 (High persistent)</th>
<th>CP Class 2 (No problem)</th>
<th>CP Class 3 (Low persistent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-persistent trajectory (27%)</td>
<td>2000</td>
<td>2.11** (0.02)</td>
<td>-0.02** (.004)</td>
<td>0.93</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>Low-persistent stable trajectory (43.3%)</td>
<td>3170</td>
<td>0.11* (0.05)</td>
<td>0.02 (0.01)</td>
<td>0.00</td>
<td>0.93</td>
<td>0.07</td>
</tr>
<tr>
<td>No-problem stable trajectory (29.5%)</td>
<td>2160</td>
<td>1.29** (0.03)</td>
<td>-0.01* (.006)</td>
<td>0.04</td>
<td>0.05</td>
<td>0.91</td>
</tr>
</tbody>
</table>

*Note.* *p<.05, **p<.001, CP= Class probabilities for the most likely latent class membership. All unweighted N rounded to 10 as per NCES requirements.
Table 2.6

**LCGA Regression Odds Ratio Results**

<table>
<thead>
<tr>
<th>Child Characteristics</th>
<th>OR 95% CI</th>
<th>OR 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>High-persistent stable</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trajectory class (HP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2.40**</td>
<td>0.48**</td>
</tr>
<tr>
<td>Black</td>
<td>2.07**</td>
<td>0.41**</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.78*</td>
<td>1.03</td>
</tr>
<tr>
<td>Other</td>
<td>0.70*</td>
<td>1.11</td>
</tr>
<tr>
<td>Birth weight in pounds</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>Reading IRT K Fall</td>
<td>1.01</td>
<td>0.98*</td>
</tr>
<tr>
<td>Math IRT K Fall</td>
<td>1.02*</td>
<td>1.01</td>
</tr>
<tr>
<td>Numbers reversed score K Fall</td>
<td>1.01</td>
<td>1.00</td>
</tr>
<tr>
<td>Attention K spring</td>
<td>0.41**</td>
<td>2.65**</td>
</tr>
<tr>
<td>Approaches to Learning K spring</td>
<td>0.60**</td>
<td>1.75**</td>
</tr>
<tr>
<td>Internalizing symptoms K spring</td>
<td>1.19</td>
<td>1.54**</td>
</tr>
<tr>
<td>Total Language Score</td>
<td>1.03</td>
<td>0.94**</td>
</tr>
<tr>
<td>IEP in Kindergarten spring</td>
<td>0.90</td>
<td>0.84</td>
</tr>
<tr>
<td>Percent of minority students in school</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td><strong>No-problem stable</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>trajectory class (NP)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Family Characteristics</th>
<th>OR 95% CI</th>
<th>OR 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socio-economic status</td>
<td>0.60**</td>
<td>1.32*</td>
</tr>
<tr>
<td>Parent Education Level</td>
<td>1.06</td>
<td>0.96</td>
</tr>
</tbody>
</table>

*Note: *p<.05, **p<.001. Reference groups are low problem class, white, and girl.
### Table 2.7

**Between Class Mean Comparison**

<table>
<thead>
<tr>
<th>Child Characteristics</th>
<th>High Persistent Group (SD)</th>
<th>No Problem Group (SD)</th>
<th>Low Persistent group (SD)</th>
<th>F or Chi square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent boys</td>
<td>74.01</td>
<td>30.52</td>
<td>51.38</td>
<td>351.23</td>
</tr>
<tr>
<td>Percent White students</td>
<td>47.87</td>
<td>55.55</td>
<td>51.52</td>
<td>6.90</td>
</tr>
<tr>
<td>Percent Black students</td>
<td>23.14</td>
<td>5.38</td>
<td>12.63</td>
<td>79.46</td>
</tr>
<tr>
<td>Percent Hispanic students</td>
<td>21.44</td>
<td>27.33</td>
<td>25.24</td>
<td>5.58</td>
</tr>
<tr>
<td>Percent Other race</td>
<td>7.52</td>
<td>11.75</td>
<td>10.61</td>
<td>6.34</td>
</tr>
<tr>
<td>Birth weight in pounds</td>
<td>6.82a (1.37)</td>
<td>6.86a (1.30)</td>
<td>6.85a (1.30)</td>
<td>1.78</td>
</tr>
<tr>
<td>Reading IRT K Fall</td>
<td>53.27b (11.08)</td>
<td>56.55b (11.86)</td>
<td>54.84c (11.07)</td>
<td>30.19</td>
</tr>
<tr>
<td>Math IRT K Fall</td>
<td>34.65c (10.49)</td>
<td>38.77b (11.86)</td>
<td>36.20c (11.07)</td>
<td>58.54</td>
</tr>
<tr>
<td>Numbers reversed K Fall</td>
<td>92.99c (16.99)</td>
<td>99.14b (16.23)</td>
<td>95.80c (16.72)</td>
<td>60.53</td>
</tr>
<tr>
<td>Attention K spring</td>
<td>2.36a (0.79)</td>
<td>3.64b (0.60)</td>
<td>3.05c (0.80)</td>
<td>918.12</td>
</tr>
<tr>
<td>Approaches to Learning K Fall</td>
<td>2.61c (0.63)</td>
<td>3.59b (0.46)</td>
<td>3.17c (0.60)</td>
<td>971.90</td>
</tr>
<tr>
<td>Total Language Score</td>
<td>18.82a (0.54)</td>
<td>18.71a (0.44)</td>
<td>18.80a (0.47)</td>
<td>1.51</td>
</tr>
<tr>
<td>Numbers reversed K Fall</td>
<td>12.85 (2.83)</td>
<td>4.71</td>
<td>8.99</td>
<td>38.23</td>
</tr>
<tr>
<td>Percent IEP in K spring</td>
<td>18.22 (27.19)</td>
<td>18.71b (17.37)</td>
<td>18.80c (22.85)</td>
<td>7.71</td>
</tr>
<tr>
<td>Externalizing in K</td>
<td>7.39a (3.96)</td>
<td>0.81b (1.35)</td>
<td>3.23c (2.61)</td>
<td>1304.33</td>
</tr>
<tr>
<td>Externalizing 1st grade</td>
<td>8.30a (3.50)</td>
<td>1.37b (1.41)</td>
<td>3.74c (2.20)</td>
<td>2537.32</td>
</tr>
<tr>
<td>Externalizing 2nd grade</td>
<td>8.48a (3.48)</td>
<td>1.20b (1.29)</td>
<td>3.73c (2.24)</td>
<td>2252.53</td>
</tr>
<tr>
<td>Externalizing 3rd grade</td>
<td>8.12a (3.60)</td>
<td>1.09b (1.28)</td>
<td>3.60c (2.23)</td>
<td>2057.55</td>
</tr>
<tr>
<td>Externalizing 4th grade</td>
<td>7.73a (3.60)</td>
<td>1.14b (1.36)</td>
<td>3.35c (1.39)</td>
<td>2058.90</td>
</tr>
<tr>
<td>Externalizing 5th grade</td>
<td>7.42a (3.78)</td>
<td>1.20b (1.39)</td>
<td>3.36c (2.48)</td>
<td>2880.04</td>
</tr>
</tbody>
</table>

**Family Characteristics**

| Socio-economic status                  | -0.22a (0.72)              | 0.09b (0.82)          | -0.02 (0.77)c             | 46.70           |
| Parent Education Level                 | 4.44a (1.68)               | 4.87b (1.94)          | 4.69b (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.
Figure 1.1. Visual representation of potential causal pathways between externalizing behavior problems and academic achievement based on Hinshaw (1992). E=Externalizing behavior problems, AA=Academic Achievement.
Figure 1.2. Mediation model between academic achievement and externalizing behavior problems.
Figure 1.3 PRISMA CHART
<table>
<thead>
<tr>
<th>Authors</th>
<th>Description of Sampling (I/12)</th>
<th>Description of Measurement (I/6)</th>
<th>Statistical Methodology, Research Design (I/14)</th>
<th>External Validity and Limitations (I/2)</th>
<th>Bias Score (I/34)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burt &amp; Roisman (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brennan et al. (2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Massetti et al. (2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McCalfe et al. (2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duncan et al. (2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breslau et al. (2009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hurry et al. (2016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morgan et al. (2018)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weeks et al. (2016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fergusson et al. (1993)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Masten et al. (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gray et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sala et al. (2016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fergusson &amp; Horwood (1995)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fitzpatrick &amp; Pagani (2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bussing et al. (2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chadwick et al. (1999)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kremer et al. (2016)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabiner et al. (2000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Realmuto et al. (2000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russel et al. (2015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Miller et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rennie et al. (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rapport et al. (1999)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Willcutt et al. (2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vaughn &amp; Haager (1994)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1.4 Risk of bias in each domain based on Quality Indicator Scores
Figure 2.1 Distribution frequencies of outcome variable at all waves
Figure 2.2 Mean teacher rated scores of externalizing behaviors from kindergarten to fifth grade
Figure 2.3. Simplified figure or proposed Latent Class Growth Model.
Figure 2.4 Mean teacher ratings of externalizing behavior through fifth grade for two and three class solutions respectively
### Appendix A

#### Quality Indicator Checklist (Tooth et al., 2005)

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Are the objectives or hypotheses of the study stated?</td>
<td>Self-explanatory.</td>
</tr>
<tr>
<td>2. Is the target population and sampling frame defined?</td>
<td>The list of units from which the study population will be drawn. Ideally, the sampling frame would be identical to the target population, but it is not always possible.</td>
</tr>
<tr>
<td>4. Is the study population defined?</td>
<td>The group selected for investigation.</td>
</tr>
<tr>
<td>5. Are the study setting (venues) and/or geographic location stated?</td>
<td>Comment required about location of research. Could include name of center, town, or district.</td>
</tr>
<tr>
<td>6. Are the dates between which the study was conducted stated or implicit?</td>
<td>Self-explanatory.</td>
</tr>
<tr>
<td>7. Are eligibility criteria stated?</td>
<td>The words “eligibility criteria” or equivalent are needed, unless the entire population is the study population.</td>
</tr>
<tr>
<td>8. Are issues of “selection in” to the study mentioned?†</td>
<td>Any aspect of recruitment or setting that results in the selective choice of participants (e.g., gender or health status influenced recruitment). Justification of number of subjects needed to detect anticipated effects. Evidence that power calculations were considered and/or conducted.</td>
</tr>
<tr>
<td>9. Is the number of participants justified?</td>
<td>Quantitative statement of numbers.</td>
</tr>
<tr>
<td>10. Are numbers meeting and not meeting the eligibility criteria stated?</td>
<td>Quantitative comparison of the different groups.</td>
</tr>
<tr>
<td>11. For those not eligible, are the reasons why stated?</td>
<td>Broad mention of the major reasons.</td>
</tr>
<tr>
<td>12. Were consenters compared with nonconsenters?</td>
<td>Total number of participants (after screening for eligibility and consent) included in the first stage of data collection. Descriptions of tools (e.g., surveys, physical examinations) and processes (e.g., face-to-face, telephone).</td>
</tr>
<tr>
<td>15. Was the number of participants at the beginning of the study stated?</td>
<td>Evidence of reproducibility of the tools used.</td>
</tr>
<tr>
<td>16. Were methods of data collection stated?</td>
<td>Evidence that the validity was examined against, or discussed in relation to, a gold standard. Confounders were defined as a variable that can cause or prevent the outcome of interest, is not an intermediate variable, and is associated with the factors under investigation.</td>
</tr>
<tr>
<td>17. Was the reliability (repeatability) of measurement methods mentioned?</td>
<td>Quantitative statement of numbers at each follow-up point.</td>
</tr>
<tr>
<td>18. Was the validity (against a “gold standard”) of measurement methods mentioned?</td>
<td>Broad mention and quantification of the major reasons.</td>
</tr>
<tr>
<td>19. Were any confounders mentioned?</td>
<td>Differences in numbers of data points (indicating missing data items) explained.</td>
</tr>
<tr>
<td>20. Was the number of participants at each stage/wave specified?</td>
<td></td>
</tr>
<tr>
<td>21. Were reasons for loss to follow-up quantified?</td>
<td></td>
</tr>
<tr>
<td>22. Was the missingness of data items at each wave mentioned?</td>
<td></td>
</tr>
</tbody>
</table>
23. Was the type of analyses conducted stated? Specific statistical methods mentioned by name.

24. Were “longitudinal” analysis methods stated? Longitudinal analyses were defined as those assessing change in outcome over two or more time points and that take into account the fact that the observations are likely to be correlated. Absolute effect was defined as the outcome of an exposure expressed, for example, as the difference between rates, proportions, or means, as opposed to the ratios of these measures.

25. Were absolute effect sizes reported?

26. Were relative effect sizes reported? Relative effects were defined as a ratio of rates, proportions, or other measures of an effect.

27. Was loss to follow-up taken into account in the analysis? Specific mention of adjusting for, or stratifying by, loss to follow-up.

28. Were confounders accounted for in analyses? Specific mention of adjusting for, or stratifying by, confounders.

29. Were missing data accounted for in the analyses? Specific mention of adjusting for, or stratifying by, or imputation of missing data items.

30. Was the impact of biases assessed qualitatively? Specific mention of bias affecting results, but magnitude not quantified.

31. Was the impact of biases estimated quantitatively? Specific mention of numerical magnitude of bias. A study is generalizable if it can produce unbiased inferences regarding a target population (beyond the subjects in the study). Discussion could include that generalizability is not possible.

32. Did authors relate results back to a target population? Discussion of generalizability beyond the target population.

33. Was there any other discussion of generalizability? Discussion of generalizability beyond the target population.

Note: † Represents selection bias at the beginning of a study. Other selection biases (i.e., loss to follow-up, missing data items) are dealt with by other checklist criteria.

Appendix B

<table>
<thead>
<tr>
<th>Search Terms</th>
<th>Low achievement</th>
<th>Statistical Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>externaliz*</td>
<td>academic fail*</td>
<td>structural Equation*</td>
</tr>
<tr>
<td>attenti* diffic*</td>
<td>underachiev*</td>
<td>path Analy*</td>
</tr>
<tr>
<td>attenti* disor*</td>
<td>academic achiev*</td>
<td>developmen* Cascade*</td>
</tr>
<tr>
<td>ADHD</td>
<td>school failure</td>
<td>longitudina*</td>
</tr>
<tr>
<td>attenti* defic*</td>
<td>read* diffic*</td>
<td>cohort stud*</td>
</tr>
<tr>
<td>opposition* Defia*</td>
<td>math diffic*</td>
<td>propensity scor*</td>
</tr>
<tr>
<td>conduct disor*</td>
<td>dyslexi*</td>
<td>causa*</td>
</tr>
<tr>
<td>conduct diffic*</td>
<td>dyscalculia</td>
<td>experimen*</td>
</tr>
<tr>
<td>aggress*</td>
<td>slow learner*</td>
<td>regression discont*</td>
</tr>
<tr>
<td>defian*</td>
<td>academic competenc*</td>
<td>simultaneous equation model*</td>
</tr>
<tr>
<td>disrupt* behavio*</td>
<td>low achiev*</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>behavio* diffic*</td>
<td>poor achiev*</td>
<td></td>
</tr>
<tr>
<td>behavio* disord*</td>
<td>low attainment</td>
<td></td>
</tr>
<tr>
<td>hyperactiv*</td>
<td>learning diffic*</td>
<td></td>
</tr>
<tr>
<td>behavi* problem*</td>
<td>learning disord*</td>
<td></td>
</tr>
<tr>
<td>problem behavior*</td>
<td>academic attainment</td>
<td></td>
</tr>
<tr>
<td>antisocial</td>
<td>academic prob*</td>
<td></td>
</tr>
<tr>
<td>inattenti*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix C

Train and Test Data Set

Latent class size and separation by Train and Test datasets

<table>
<thead>
<tr>
<th>Class</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training dataset</td>
<td>Test dataset</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>980</td>
<td>26.78</td>
<td>970</td>
<td>26.38</td>
</tr>
<tr>
<td>2</td>
<td>1060</td>
<td>28.82</td>
<td>1070</td>
<td>29.17</td>
</tr>
<tr>
<td>3</td>
<td>1630</td>
<td>44.39</td>
<td>1630</td>
<td>44.45</td>
</tr>
</tbody>
</table>

Classification probabilities for most likely latent class membership

<table>
<thead>
<tr>
<th>Class</th>
<th>Training dataset</th>
<th>Test dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0.927</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>0.00</td>
<td>0.932</td>
</tr>
<tr>
<td>3</td>
<td>0.036</td>
<td>0.049</td>
</tr>
<tr>
<td>Entropy</td>
<td>0.827</td>
<td></td>
</tr>
</tbody>
</table>