Developmental Psychology
2010, Vol. 46, No. 5, 976–983

0012-1649/10/$12.00 DOI: 10.1037/a0018878

Early Behavioral Associations of Achievement Trajectories

Kevin J. Grimm and Joel S. Steele
University of California, Davis

Andrew J. Mashburn
University of Virginia

Margaret Burchinal
University of North Carolina, Chapel Hill

Robert C. Pianta
University of Virginia

Duncan et al. (2007) examined associations between early behavioral and cognitive skills with later achievement. These associations were examined in 6 different data sets and results converged to suggest that early behavioral competences or problems had little, if any, prediction to later achievement and that attentional competences had small positive relations with later achievement. In contrast, cognitive abilities were by far the strongest predictors of achievement. We provide and investigate potential reasons why Duncan et al. found little to no association between behavior and later achievement in a reanalysis of data from 3 studies previously analyzed by Duncan et al. Potential reasons include the validity of the behavioral measures, treatment of the behavioral measures as continuous as opposed to categorical, and the choice of data analytic method. In this article, we discuss these issues at greater length and address them in our reanalysis. We also bring into question the nature of the relationship between behavior and achievement. Generally, our reanalysis supports the idea that attention measures are more predictive than behavioral measures; however, certain behavior measures showed small to moderate associations to concurrent levels of academic achievement and changes in academic achievement through elementary school.

Keywords: achievement, growth, behavior, attention, longitudinal

Duncan et al. (2007) examined associations between early behavioral and cognitive skills and with later achievement. These associations were examined in six different data sets, and the results converged to suggest that (a) early cognitive measures were the most predictive of later reading and mathematics achievement; (b) early behavior measures, such as externalizing and internalizing behavior, did not show much, if any, predictive power of later achievement; and (c) early measures of attention had small positive associations with later achievement. These results were interpreted not only in relation to developmental theory but also in light of the appropriate focus of early education programs that aim to improve school-age achievement.

There are several potential reasons for the lack of an association between early behavior indicators and later achievement. These include the validity of behavioral measures, the focus on continuous relations between behavioral indicators and later achievement (in contrast to group differences), and the relationship between early behavior and achievement (i.e., there may not be an association between early behavior and later achievement). In addition, Duncan et al. (2007) chose to use data from a few of the available time points (i.e., early measures collected at school entry [∼ age 5] and later achievement measured between third and eighth grade [∼ age 13 to 14]) rather than using the available data between school entry and eighth grade, which would increase the reliability of the outcome (i.e., achievement change). In this article, we discuss these issues at greater length and address several issues in a reanalysis of three of the data sets used by Duncan et al.

Sources of Variation in Behavioral Measures

The behavioral measures used in the studies were considered to be reliable, usually measured by coefficient alpha. In fact, high levels of reported internal consistency were used as an argument to support the validity of Duncan et al.'s (2007) results. However, informant-based behavioral measures are indirect assessments, as they rely on teacher and/or parental reports, and have been shown to be subject to informant bias. For example, several multitrait–multimethod analyses of informant-based behavior rating scales have shown substantial amounts of informant-based variance. In a recent study (Konold & Pianta, 2007) of first grade students, the Child Behavior Checklist was completed by the study child's mother, father, and teacher. The ratings of child behavior were shown to be heavily influenced by the informant. Often more informant-based variance than trait-based variance was reported.

Kevin J. Grimm and Joel S. Steele, Department of Psychology, University of California, Davis; Andrew J. Mashburn and Robert C. Pianta, Center for the Advanced Study of Teaching and Learning, University of Virginia; Margaret Burchinal, Frank Porter Graham Child Development Institute, University of North Carolina at Chapel Hill.

This research was supported by the National Center for Research on Early Childhood Education, Institute of Education Sciences, U.S. Department of Education (Grant R305A06021) and a National Science Foundation REECE Program Grant (Grant DRL-0815787). The opinions and views expressed in this article are those of the authors and do not necessarily represent the views and opinions of the U.S. Department of Education.

Correspondence concerning this publication should be addressed to Kevin J. Grimm, Psychology Department, University of California, Davis, One Shields Avenue, Davis, CA 95616. E-mail: kjgrimm@ucdavis.edu

Therefore, in many cases rating scales provide more information about the informant who completed the form than the child. Konold and Pianta's (2007) results regarding the informant issues in child behavior rating scales were in line with previously published research on the topic (e.g., Byrne & Bazana, 1996; Byrne & Schneider, 1986; Cole, 1990; Epkins & Meyers, 1994; Greenbaum, Decrick, Prange, & Friedman, 1994; Hong, Paunonen, & Slade, 2008; Matson & Nieminen, 1987; Wolfe et al., 1987).

In addition, Grimm and Pianta (2009) conducted a multitrait–multimethod analysis with data from the Child Behavior Checklist to examine predictors of trait and informant variance in externalizing and internalizing behaviors. Predictors included demographic measures and measures of the mother's and father's well-being. Maternal depression was shown to have a greater association with the maternal informant factor than any of the traits. Similarly, paternal depression was found to have a greater relationship with the paternal informant factor than any of the traits. This set of research demonstrates that even though behavior rating scales have high internal consistency, their validity is limited because such a large portion of observed variation is attributable to the informant.

Continuous Associations Versus Group Differences

Measures of social behavior, particularly problem behavior, tend to have nonnormal distributions. In all studies of typical samples, distributions for problem behaviors tend to be positively skewed with a preponderance of zeroes, indicating that a large percentage of children do not have any behavior problems. For example, in the NICHD Study of Early Child Care (NICHD Early Child Care Research Network [NICHD-ECCRN], 1997), the teacher-reported measure of aggression in kindergarten has a skew of 2.4, a minimum score of 0, a maximum score of 49, a mode of 0, a first quartile of 0, a median of 1, and a mean of 4.4. Thus, more than half of the sample has a score of 0 or 1. There is little doubt that problem behavior is better characterized as categorical than continuous.

Duncan et al. (2007) analyzed the predictive effects of problem behaviors in terms of linear associations, treating the behavior problem variables as continuous, and this may not be reasonable (e.g., outliers have large influence). However, Duncan et al. did evaluate nonlinear effects using spline regression and reported no important differences. Another way to examine nonlinear effects is through group differences. Behavior rating scales are often designed to identify children who may have clinical or significant levels of behavior problems and are not designed to discriminate children with low or even modest levels of behavior problems. Therefore, it may be better to categorize children on the basis of their level of behavior problems rather than assume a continuous meaningful distribution. That is, it may be important to distinguish between children with low levels of behavior problems, children with moderate levels of behavior problems, and children with high levels of behavior problems. It may be that there are no important differences between children with low and moderate levels of behavior problems, but significant differences may exist between children who have low and high levels of behavior problems. In support of this notion, several investigators have used latent class analysis to examine groupings of children based on behavior rating scales (e.g., Hill, Degnan, Calkins, & Keane, 2006).

Associations Between Behavior and Achievement

Children's behavior and achievement have been shown to be associated in several articles (e.g., Arnold, 1997; Bub, McCartney, & Willett, 2007; McClelland, Morrison, & Holmes, 2000), but the association is too often examined in cross-section; children who have more behavior problems tend to also have lower achievement, concurrently. Thus, the direction of this effect is unknown. Is it that low achievement leads to frustration, which leads to behavior problems, or is it that behavior problems lead to low attention, which in turn leads to low achievement? Longitudinal investigations of this association are mixed. For example McArdle and Hamagami (2001) examined lead–lag associations between behavior problems and reading achievement with longitudinal data from the National Longitudinal Survey of Youth. McArdle and Hamagami (2001) found that low reading achievement led to greater increases in behavior problems and not the opposite, which lends support for low achievement as a precursor to behavior issues. Similarly, Grimm (2007) examined the time-dependent relationships between children's depression and achievement and found a small lead–lag relationship, with low achievement leading to increases in children's level of depression. However, Stuhlman et al. (2009) fit longitudinal models to achievement and behavior data from the NICHD Study of Early Child Care and Youth Development (SECCYD) and found that academic skills led to changes in behavior and that behavior led to changes in academic skills.

Another important point surrounding the association between achievement and behavior is timing. Most of the previously discussed research examined school-age children. However, this association is likely to be present in children who have yet to enter school. From birth, children are constantly exposed to stimuli that help to develop their cognitive capabilities, and academic or cognitive variation can be observed in infancy. In addition, variation in children's temperament and behavior can also be observed at very young ages. This adds to the difficulty of determining the time dependency between these constructs and presents the idea that there may not be a strong link between behavior and achievement once school starts, but that is not to say that this relationship is not present in younger children.

Longitudinal Data and Models

Duncan et al. (2007) fit two-occasion autoregression models with covariates. In this type of model, later achievement scores are predicted by previous achievement scores and the set of covariates, which included measures of behavior, cognition, and attention. This type of model is appropriate to study change with two-occasion longitudinal data; however, many of the data sets analyzed by Duncan et al. (2007) had at least four occasions in which achievement, the major construct of interest, was measured. Additional measurement occasions allow for the study of change and the separation of measurement error (see McArdle & Epstein, 1987; Meredith & Tisak, 1990; Rogosa & Willett, 1985). The growth curve model is a more powerful analytic tool for studying change with several measurement occasions. Concurrent associations between achievement and the set of covariates can also be analyzed in this framework.

Current Project

In the current project we used growth curve analysis to examine the associations between early measures of behavior and attention with changes in achievement through elementary school. We analyzed data from the NICHD SECCYD (NICHD-ECCRN, 1997), the Early Childhood Longitudinal Study Kindergarten Cohort (ECLS-K; National Center for Education Statistics, 2001), and the National Longitudinal Survey of Youth–Children and Young Adults (NLSY-CYA; Center for Human Resource Research, 2004) to evaluate these associations. In addition, we created trichotomies from each behavioral and attention measure to examine differences between children who have low (normative) levels of behavior or attention problems and children with moderate (problematic) and high (clinical) levels of behavior or attention issues. The results are described using effect sizes (Cohen's *d*) for a straightforward comparison of results across predictors and studies.

Method

Data

We analyzed data from three of the longitudinal studies previously analyzed by Duncan et al. (2007). These data sets included the SECCYD, the ECLS-K, and the NLSY-CYA. Descriptions of these studies appear in several publications (e.g., Duncan et al., 2007). In the following brief descriptions we discuss the participants and focus on the variables analyzed in the current project.

SECCYD. The SECCYD (NICHD-ECCRN, 2000, 2001, 2002) provides detailed, repeated, and comprehensive assessments of family, child care, and schooling context (observations) as well as standardized assessments of child outcomes in multiple domains including language, literacy, social development, and health for 1,364 children from 10 sites across the country. The measures of academic achievement examined here include the repeated assessments of two Woodcock-Johnson-Revised subtests: Letter–Word Identification and Applied Problems. These subtests were administered when the children were 54 months old and in the spring of the first, third, and fifth grade. The set of covariates include measures of early behavior problems (i.e., mother-reported internalizing and externalizing) measured at 36 months and attention (i.e., teacher-reported attention problems and the continuous performance task) as well as children's behavior in the fall of kindergarten (school entry). These measures included teacher-reported academic competence and positive social skills from the Social Skills Rating Scale and externalizing and internalizing behaviors from the Teacher Report Form of the Child Behavior Checklist.

ECLS-K. The ECLS-K (National Center for Education Statistics, 2001) is a nationally representative sample of 21,260 children who attended kindergarten in 1998–1999. These children were followed from the fall of kindergarten to the end of eighth grade. Currently, data from the fall of kindergarten through the end of fifth grade are publicly available. Data come from direct assessments and teacher, parent, and school reports. Reading and mathematics achievement tests were administered in the fall and spring of kindergarten and first grade and in the spring of third and fifth grade. These achievement scores were the vertically linked item response theory scaled scores, which are appropriate for examining change. Behavior (i.e., internalizing, externalizing, interpersonal skills, and self control) and attention skills (i.e., approaches to learning) were measured in the fall of kindergarten with teacher-reports.

NLSY-CYA. The NLSY-1979, initiated in 1979, was a multistage stratified random sample of 12,686 participants between the ages of 14 and 21 (Center for Human Resource Research, 2004). African American, Hispanic, and low-income youths were oversampled. In 1986, children of NLSY-1979 female participants were measured; this sample, which was assessed every 2 years through 2006, makes up the NLSY-CYA. The NLSY-CYA contains measures of academic achievement (i.e., Peabody Individual Achievement Tests: Reading Comprehension and Mathematics), parental reports of children's behavior based on the Behavior Problems Index (Antisocial Behaviors, Hyperactivity, Headstrong, and Anxious/Depressed) and the Temperament Scale (Compliance and Sociability). The sample used in the present analysis included approximately 9,000 children whose academic achievement was measured between ages 5 and 14. (SAS scripts for combining each of these data sets are available online at http://psychology.ucdavis.edu/labs/Grimm/personal/downloads.html)

Analytic Techniques

Categorizations of behavior scales. Children were categorized into *normative*, *problematic*, and *clinical* groups for each behavior measure. Different behavior scales were used in the SECCYD, NLSY, and ECLS-K; the measures used in the NLSY and ECLS-K did not have set cutoff points for problematic and clinical levels of behavior problems. Thus, categorizations for these scales were based on the observed distributions in the data. For negative behavior traits (e.g., externalizing and internalizing) where higher scores indicate more problems, the normative group had scores lower than the 1 standard deviation above the mean, the problematic group had scores between 1 and 1.5 standard deviations above the mean, and the clinical group had scores above 1.5 standard deviations above the mean. For positive behavior traits (e.g., social skills and attention), the groups were made in the opposite direction. For the SECCYD, a *t* score above 60 was considered problematic and a *t* score above 65 was considered clinical, which related to 1 and 1.5 standard deviations above the mean, respectively.

Growth curve analysis. Growth curve analysis (McArdle & Epstein, 1987; Meredith & Tisak, 1990; Rogosa & Willett, 1985) is an analytic technique for modeling systematic within-person change across a series of repeated measurements and between-person differences in those changes. Given repeated measurement of a variable, Y, for $n = 1$ to N participants on $t = 1$ to T occasions (or ages), a growth curve with an intercept (g_{0n}) and one slope (g_{1n}) can be written as

$$Y[t]_n = g_{0n} + g_{1n} \cdot A_1[t] + e[t]_n, \quad (1)$$

where g_{0n} is the intercept for subject n, g_{1n} is a slope for subject n, A_1 is a vector of basis coefficients indicating the relationship between the slope and the observed scores, and $e[t]_n$ is a time-dependent residual that is uncorrelated with the intercept and slope. The intercept and slope are assumed to follow multivariate normal distributions described by means, variances, and covariance(s). The basis coefficients can be fixed to test specific hypoth-

eses, such as linear ($A_1[t] = t$) or exponential [$A_1[t] = \exp(\alpha \times t)$] growth, or can be estimated with minimal identification constraints in a latent basis growth model (e.g., Meredith & Tisak, 1990). The time-dependent residuals were assumed to have a mean of zero and variances and were unrelated to other variables and each other. The growth curve is flexible as additional slopes can be added to examine nonlinear developmental patterns and/or the basis coefficients can follow nonlinear patterns to nonlinear change (see Grimm & Ram, 2009; Ram & Grimm, 2007).

Time-invariant covariates were included in the growth curve as predictors of the intercept and slope in regression-like equations, which can be written as

$$g_{0n} = b_{00} + b_{10} \cdot x_{1n} + \ \ldots \ + b_{k0} \cdot x_{kn} + d_{0n}$$

$$g_{1n} = b_{01} + b_{11} \cdot x_{1n} + \ \ldots \ + b_{k1} \cdot x_{kn} + d_{1n}, \qquad (2)$$

where g_{0n} and g_{1n} are the Level 1 intercept and slope from Equation 1; b_{00} is the Level 2 intercept term for the Level 1 intercept; $b_{10} - b_{k0}$ are regression coefficients for the time-invariant covariates ($x_{1n} - x_{kn}$) for the Level 1 intercept or the effects the covariates have in helping explain the between-person variation in the intercept; d_{0n} is the Level 2 disturbance term for the intercept; b_{01} is the Level 2 intercept term for the Level 1 slope; $b_{11} - b_{k1}$ are regression coefficients for the time-invariant covariates for the Level 1 slope; and d_{1n} is the Level 2 disturbance term for the slope.

Combined item response and growth curve model. For the SECCYD we fit a combined item response and growth curve model (see McArdle, Grimm, Hamagami, Bowles, & Meredith, 2009). The combined item response growth curve model was fitted to account for differential reliability, as the Woodcock-Johnson was not well targeted for children with low levels of ability at 54 months. The item response part of model can be written as

$$P(z[t]_{in} = 1|\theta[t]_n, \alpha_i, \beta_i) = \frac{\exp[\alpha_i(\theta[t]_n - \beta_i)]}{1 + \exp[\alpha_i(\theta[t]_n - \beta_i)]}, \qquad (3)$$

where $P(z[t]_{in} = 1|\theta[t]_n, \alpha_i, \beta_i)$ is the probability of a correct response to the ith item by person n at time t, conditional on the person's level of proficiency at time t($\theta[t]_n$) and the item's discrimination (α_i) and difficulty (β_i) parameters. It is important to note that item discrimination and difficulty were assumed to be invariant across time (i.e., longitudinal measurement equivalence), whereas the child's latent trait was time-dependent. This item response model is a longitudinal extension of the two-parameter logistic model. The Woodcock-Johnson tests are Rasch scaled, so the discrimination parameter was fixed at 1, reducing the model to a longitudinal one-parameter logistic (Rasch, 1960). In the combined item response growth model, $\theta[t]_n$ was assumed to follow the growth curve model of Equation 1 ($\theta[t]_n$ replaces $Y[t]_n$) and was extended to include covariates as in Equation 2.

Incomplete data. Incomplete data were present in all studies. In the SECCYD, incomplete data in the set of covariates were imputed using information from demographic and early behavior problems variables. Teacher-reported behavior problems were approximately 15% incomplete, and mother-reported behavior problems were approximately 5% incomplete. Imputations were done using the continuous forms of the variables when possible (and then categorized for analysis). Ten imputations were conducted; between-imputation variance was generally small, so data from the first imputation were used in the analysis. Incomplete data in the outcome measures were not imputed, as the models were fit to the available data.

Incomplete data in the NLSY-CYA were apparent for several reasons, mostly dealing with the timing of measurement occasions. In 1986, children varied in age and some children who would be assessed later were not yet born. School entry covariates were incomplete for 28% to 30% of the participants, depending on the specific covariate. Longitudinal reading and mathematics measures were largely incomplete due to the measurement schedule. Incomplete data were also apparent in the ECLS-K. School entry covariates were incomplete for 6% to 14% of the participants, depending on the specific covariate. Longitudinal reading and mathematics measures were also incomplete. Most notably, only 30% of participants were tested in the fall of kindergarten because of the study design. Attrition was evident over the course of the study; 64% of the total sample was tested in fifth grade compared with 95% and 81% in the spring of first and third grade, respectively.

In the NLSY-CYA and ECLS-K, full information maximum likelihood (FIML) estimation (Little & Rubin, 1987) was used to analyze all available data. The use of FIML and imputation assumes the incomplete data are missing at random (MAR); the missing data mechanism is related to the measured data included in the model. Growth curve models for the ECLS-K and NLSY-CYA data were fit using Mplus (Muthén & Muthén, 2008); the combined item response growth curve models fit to the SECCYD data were fit using WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000).

Results

The results are presented in six sections. First, we provide an overview of the growth modeling results and the similarity of these results across studies. Next, we describe concurrent and longitudinal associations between achievement and demographic measures, attention, internalizing, externalizing, and social skills. We focus on the similarities and differences in the results across studies and report average effect sizes. Effect sizes (Cohen's d) for all covariates included in the models are contained in Tables 1, 2, and 3 for the SECCYD, ECLS-K, and NLSY-CYA, respectively.

Growth Modeling Results

The latent basis growth model, in which the first and last basis coefficients ($A[t]$) were fixed at 0 and 1, respectively, and the remaining coefficients were estimated, was the best fitting model; however, the structure of these changes differed depending on the outcome measure. The structure of changes for the Letter–Word Identification and Applied Problems (from the SECCYD) as well as the PIAT Reading Comprehension and Mathematics (from the NLSY-CYA) were similar; the rate of change was greatest during early childhood and gradually declined as children progressed through school, approximating exponential growth with a negative rate of change. This was different from how changes occurred for reading and mathematics in the ECLS-K, where the rate of change was slow during kindergarten, greatest from first through third grades, and slow from third through fifth grade. In all cases, there was significant variation in the intercept, centered around school entry (i.e., 54 months in SECCYD, fall of kindergarten in ECLS-K, and age 5 in NLSY-CYA), and the slope, which repre-

Table 1
Effect Sizes (Cohen's d) for Demographic Characteristics and Behavioral Effects on the Intercept and Slope for Letter–Word Identification and Applied Problems From the NICHD Study of Early Child Care and Youth Development

	Letter–word identification		Applied problems	
Covariate	Intercept	Slope	Intercept	Slope
Demographics				
Female	0.23*	−0.25*	0.17*	−0.23*
African American	−0.31*	−0.52*	−0.73*	−0.03
Hispanic	−0.37*	0.36	−0.41*	0.28
Poverty	−0.27*	0.21	−0.21*	0.18
Maternal education				
High school	0.30*	0.23	0.27*	0.14
College degree	0.75*	0.06	0.62*	0.20
Mother's age at birth				
<18	0.47*	−0.04	0.09	0.20
>35	0.01	0.01	0.17	−0.12
Special ed.	0.02	−0.08	−0.07	0.39
Teacher-reported behavior				
Academic competence				
Problematic	−0.68*	−0.24	−0.62*	0.02
Clinical	−0.27	−0.11	−0.96*	−0.14
Social skills				
Problematic	−0.09	−0.09	−0.12	0.00
Clinical	0.25	−0.08	−0.46*	0.00
Attention problems				
Problematic	−0.38*	0.55*	−0.33*	0.21
Clinical	−0.62*	0.31	−0.11	−0.28
Internalizing				
Problematic	−0.34*	0.16	−0.23*	0.22
Clinical	−0.35*	0.23	0.09	0.12
Externalizing				
Problematic	0.17	0.17	0.33*	−0.25
Clinical	0.22	−0.27	0.09	0.15
Mother-reported behavior (age 3)				
Externalizing				
Problematic	−0.03	0.05	0.18	−0.09
Clinical	0.09	−0.03	−0.10	−0.12
Internalizing				
Problematic	−0.11	−0.03	−0.13	0.09
Clinical	−0.33*	−0.05	−0.24	0.11
Attention task				
Continuous performance task				
Problematic	−0.33*	0.15	−0.32*	0.17
Clinical	−0.18*	0.23	−0.55*	0.57*

* $p < .05$.

sented the total amount of change from school entry to the end of elementary school (i.e., spring of fifth grade in SECCYD & ECLS-K, age 14 in the NLSY-CYA). Finally, there was a positive correlation between the intercept and slope for reading and mathematics measures in the ECLS-K and the NLSY-CYA, but a negative intercept–slope correlation for the Woodcock-Johnson measures in the SECCYD.

Associations Between Demographic Measures and Achievement

Several common demographic measures were included as control variables in the prediction of achievement at school entry and achievement change during elementary school. There were several consistent effects across studies; however, some effects were study specific. In most cases, girls outperformed boys at school entry (average $d = 0.12$); however, boys tended to show more growth during elementary school (average $d = -0.18$). African American and Hispanic children, compared with European American and Asian American children, tended to have lower levels of achievement at school entry (average $d = -0.31$ and $d = -0.34$ for African American and Hispanic children, respectively). In addition, African American children tended to show less growth during school (average $d = -0.34$), whereas Hispanic children showed growth similar to European American and Asian American students (average $d = -0.02$). Achievement gaps were also evident for children living in poverty, as there was a poverty-based achievement gap at school entry in all studies (average $d = -0.23$), and most studies showed a widening of this achievement gap as children progressed through school (average $d = -0.04$). Maternal education had consistent positive effects on achievement at school entry (average $d = 0.29$ and average $d = 0.74$ for high school education and college education, respectively) and had positive effects on changes in achievement (average $d = 0.28$ and average $d = 0.48$ for high school education and college education, respectively).

Associations Between Attention and Achievement

Attention was measured somewhat differently in each study (Continuous Performance Task and Teacher-Reported Attention

Table 2
Effect Sizes (Cohen's d) for Demographic Characteristics and Behavioral Effects on the Intercept and Slope for Mathematics and Reading From the Early Childhood Longitudinal Study Kindergarten Cohort

	Mathematics		Reading	
Covariate	Intercept	Slope	Intercept	Slope
Demographics				
Female	−0.11**	−0.30**	0.07**	0.06**
African American	−0.35**	−0.64**	−0.13**	−0.54**
Hispanic	−0.40**	−0.11**	−0.30**	−0.17**
Poverty	−0.18**	−0.15**	−0.18**	−0.17**
Maternal education				
High school	0.31**	0.33**	0.34**	0.43**
College degree	0.86**	0.69**	0.86**	0.81**
Special ed.	−0.35**	−0.43**	−0.28**	−0.54**
Behavior				
Self-control				
Problematic	0.00	0.05	0.05	0.04
Clinical	0.07	−0.03	0.08*	−0.04
Externalizing				
Problematic	−0.08	−0.06	−0.09*	−0.02
Clinical	−0.06	−0.08*	−0.05	−0.06
Internalizing				
Problematic	−0.09**	−0.01	−0.08*	−0.01
Clinical	−0.13**	−0.12**	−0.09**	−0.09*
Interpersonal skills				
Problematic	−0.06	−0.04	−0.05	−0.08*
Clinical	−0.08*	0.01	−0.07*	−0.07
Attention				
Problematic	−0.48**	−0.45**	−0.41**	−0.39**
Clinical	−0.66**	−0.76**	−0.48**	−0.60**

* $p < .05$. ** $p < .01$.

Table 3
Effect Sizes (Cohen's d) for Demographic Characteristics and Behavioral Effects on the Intercept and Slope for Reading Comprehension and Mathematics From the National Longitudinal Survey of Youth: Children and Young Adults

	Reading comprehension		Mathematics	
Covariate	Intercept	Slope	Intercept	Slope
Demographics				
Female	0.22**	−0.11**	0.08*	−0.27**
African American	0.14**	−0.71**	−0.47**	−0.50**
Hispanic	−0.20**	−0.21**	−0.38**	−0.28**
Poverty	−0.26**	−0.15**	−0.25**	−0.15**
Maternal education				
High school	0.29**	0.25**	0.25**	0.29**
College degree	0.66**	0.53**	0.70**	0.59**
Special needs	−0.52**	−0.73**	−0.47**	−0.72**
Behavior				
Hyperactivity				
Problematic	−0.08	−0.05	−0.12*	−0.06
Clinical	−0.23**	−0.03	−0.20*	−0.21**
Antisocial Behavior				
Problematic	0.04	−0.16**	0.06	−0.03
Clinical	0.03	−0.28**	0.01	−0.14*
Anxious/Depressed				
Problematic	0.04	0.07	−0.01	0.00
Clinical	0.01	0.07	0.10	−0.04
Headstrong				
Problematic	−0.16*	0.05	−0.04	−0.02
Clinical	−0.07	−0.01	−0.02	0.00
Compliance				
Problematic	−0.10	−0.12*	−0.14*	0.04
Clinical	−0.09	−0.13*	−0.16*	−0.06
Sociability				
Problematic	−0.19**	−0.10**	−0.23**	−0.00
Clinical	−0.59**	−0.12	−0.46**	−0.22**

* $p < .05$. ** $p < .01$.

problems in SECCYD; Teacher-Reported Approaches to Learning in ECLS-K; and Parent-Reported Hyperactivity in the NLSY-CYA) but resulted in some of the larger effects, which were mostly consistent across studies. Specifically, clinical levels of attention problems were negatively related to concurrent levels of reading and mathematics achievement in all studies. Problematic levels of attention problems were negatively associated with concurrent reading achievement in the SECCYD and ECLS-K and with concurrent mathematics achievement in all studies. Average effect sizes were $d = -0.31$ and $d = -0.38$ for problematic and clinical levels of attention problems with concurrent levels of achievement.

Associations between attention and changes in achievement were study dependent. For example, problematic teacher-rated attention was positively related to changes in reading achievement, and clinical levels of attention problems, based on the continuous performance task, were related to greater changes in mathematics in the SECCYD. Clinical and problematic levels of attention problems were negatively associated with changes in achievement with moderate to large effect sizes (average $d = -0.55$) in the ECLS-K. Finally, clinical levels of parent-rated hyperactivity were negatively associated with mathematics change in the NLSY-CYA. Average effect sizes for problematic and clinical attention problems were $d = 0.02$ and $d = -0.10$ across the three studies.

Associations Between Internalizing Behavior and Achievement

Overall, internalizing behaviors had a small negative association with concurrent levels of achievement and little association with changes in achievement. For example, problematic and clinical levels of internalizing behaviors had small negative relations to achievement in the fall of kindergarten in the ECLS-K; teacher-reported internalizing behaviors were negatively predictive of reading and mathematics achievement in the SECCYD; however, parent-reported anxious and depressed behaviors were not predictive of reading or mathematics achievement at school entry in the NLSY-CYA. Average effect sizes were $d = -0.12$ for the associations between problematic and clinical internalizing behaviors with achievement at school entry.

Significant associations between internalizing behaviors and changes in achievement were only found for the ECLS-K. These effects were small ($d = -0.12$, $d = 0.09$) and only shown for children with problematic levels of internalizing behaviors. Average effect sizes for were $d = 0.06$ and $d = 0.03$ for problematic and clinical levels, respectively.

Associations Between Externalizing Behavior and Achievement

Associations between externalizing behaviors and achievement at school entry showed some expected and unexpected patterns. In the ECLS-K, problematic levels of externalizing behaviors were negatively associated with reading achievement at school entry; however, problematic externalizing behaviors were positively associated with mathematics achievement in the SECCYD. In both cases, effect sizes were small. Average effect sizes were $d = 0.07$ and 0.03 for the association between problematic and clinical levels of externalizing behaviors with achievement at school entry, respectively.

Associations between externalizing behaviors at school entry and changes in achievement were also mixed. Problematic and clinical levels of antisocial behavior had negative associations with changes in achievement in the NLSY-CYA and clinical levels of externalizing behaviors had negative association with changes in mathematics in the NLSY-CYA and the ECLS-K. Average effect sizes between externalizing behaviors and achievement change were $d = -0.05$ and -0.10 for problematic and clinical levels, respectively.

Associations Between Social Skills and Achievement

Associations between social skills (social skills in the SECCYD, interpersonal skills in the ECLS-K, and sociability in the NLSY-CYA) and achievement at school entry were more consistent across studies than internalizing and externalizing behaviors. Significant associations were shown for problematic and clinical levels of sociability in the NLSY-CYA; clinical levels of interpersonal skills had small negative associations with achievement in the ECLS-K; and clinical levels of social skills had a moderate negative association with mathematics achievement in the SECCYD. Average effect sizes between social skills and achievement were $d = -0.12$ and $d = -0.24$ for problematic and clinical levels, respectively. Relations between socials skills and changes

in achievement were small and mostly nonsignificant. Average effect sizes were $d = -0.05$ and -0.08 for the association between problematic and clinical levels of social skills with changes in achievement, respectively.

Discussion

Summary of Findings

The approach taken in this project resulted in a series of analyses focused on concurrent and longitudinal associations between behavior and academic achievement. In addition, we focused on comparing children with normative levels of behavior and attention problems with children who displayed problematic or clinical levels to allow for nonlinear effects and determine whether severe behavior issues were necessary before behavior would manifest its impact on children's academic development. Even though we took a different approach from Duncan et al. (2007), our findings generally support their conclusion that attention skills were more important than children's behavior in predicting academic achievement and changes in academic achievement. Behavior problems generally had small and often nonsignificant effects on changes in achievement through elementary school. However, there were several exceptions and interesting caveats. These exceptions and caveats were often study specific and may represent differences in the outcome measures, covariates included in the model, and/or sample populations.

Problems with social skills had sizable negative associations with concurrent mathematics ability in the SECCYD and with achievement in the NLSY. Children with clinical levels of social skills problems scored half a standard deviation below children considered to have normative levels of social skills. Internalizing behaviors showed small to moderate negative concurrent associations with reading and mathematics in the SECCYD and ECLS-K. Finally, externalizing behaviors had, if anything, positive concurrent associations with achievement. This finding was restricted to the SECCYD sample, however, and although striking, it was not entirely unexpected; Duncan et al. (2007) found a similar association for children in the SECCYD, as there were sizable positive effects of externalizing behavior for teacher ratings of achievement and reading achievement test scores. Beyond the SECCYD, this positive link between achievement and externalizing behaviors has been reported by Dmitrieva, Steinberg, and Belsky (2007) in an analysis of participants from the ECLS-K.

As noted, attention effects were generally greater than behavior effects. This conclusion was supported by analyses of the ECLS-K and the SECCYD studies, but this was not the case for the NLSY-CYA. The lack of an attention effect in the NLSY-CYA was likely related to how attention was measured: parent-rated hyperactivity as opposed to teacher-rated attention skills in the SECCYD or teacher-rated approaches to learning in the ECLS-K.

Comparison of Effect Sizes

The findings regarding the associations between early child behavior and attention with early academic and changes in academic skills through elementary school were mixed. There were only few associations that cut across all three studies; however, in each study there were several significant associations, and a few of the associations represented moderate effect sizes. Another way to put the effect sizes in context is to compare them with effect sizes for child characteristics, because researchers are likely to be more familiar with differences in children's academic skills related to poverty and maternal education.

In the NICHD SECCYD study, concurrent associations between teacher-rated attention problems and early reading skills were approximately the same size as associations between maternal education and early reading skills. Associations for internalizing behavior and early achievement were slightly larger than the poverty effect for early reading and mathematics skills in the NICHD. For the ECLS-K, reading and mathematics differences based on children's attention problems were on par with reading and mathematics differences based on special education status and maternal education. In the NLSY, reading and mathematics differences based on the child's sociability were on par with differences based on special education status, maternal education, and poverty status. Furthermore, differences in achievement change based on the amount of antisocial behavior problems were on par with differences based on poverty status.

Concluding Remarks

The results described here were generally in line with those described by Duncan et al. (2007), even though growth models were fit and the behavior and attention measures were used to categorize children into normative, problematic, and clinical groups. Overall, the behavioral and attention associations with concurrent levels of academic achievement and changes in academic achievement were study dependent, even though moderate effects were shown for several behavior constructs within each study. Thus, behavior problems were related to academic achievement but which behavior problems were related to academic achievement were study dependent, and the majority of effects were concurrent. Differences in results between samples may be due to study design features, but there was evidence that behavior problems were related to academic achievement.

Another issue is how gender, ethnicity, and poverty status were considered in our analyses and the analyses presented by Duncan et al. (2007). Only main effects were examined, as we did not consider gender, ethnicity, or poverty status as moderating variables of the effects of early behavior and attention. As such, we assumed the associations between early behavior and attention with achievement and changes in achievement did not depend on gender, ethnicity, or poverty status. Obviously, this assumption may not hold true, and there are likely to be some significant interactions; however, these effects are likely to be small and study specific and were not pursued here.

A final point to consider is that several measures of behavior problems were included in these analyses, and these behavior problems were likely to co-occur, especially because informant-based behavioral rating scales tend to have sizable amounts of informant-based variance. The effects reported here are additive, and having several behavior problems would result in sizable differences in achievement. In addition, for a few behavior and attention measures there were both concurrent and longitudinal effects, meaning that the initial performance gap at school entry increases as the children progress through school, resulting in sizable effects at the end of elementary school.

References

Arnold, D. H. (1997). Co-occurrence of externalizing behavior problems and emergent academic difficulties in high-risk boys: A preliminary evaluation of patterns and mechanisms. *Applied Developmental Psychology, 18,* 317–330.

Bub, K. L., McCartney, K., & Willett, J. B. (2007). Behavior problem trajectories and first-grade cognitive ability and achievement skills: A latent growth curve analysis. *Journal of Educational Psychology, 99,* 653–670.

Byrne, B. M., & Bazana, P. G. (1996). Investigating the measurement of social and academic competencies for early/late preadolescents and adolescents: A multitrait–multimethod analysis. *Applied Measurement in Education, 9,* 113–132.

Byrne, B. M., & Schneider, B. H. (1986). Student–teacher concordance on dimensions of student social competence: A multitrait–multimethod analysis. *Psychopathology and Behavioral Assessment, 8,* 263–279.

Center for Human Resource Research. (2004). *A guide to the 1979–2002 National Longitudinal Survey of Youth.* Washington, DC: U.S. Department of Labor, Bureau of Labor Statistics.

Cole, D. A. (1990). Relation of social and academic competence to depressive symptoms in childhood. *Journal of Abnormal Psychology, 99,* 422–429.

Dmitrieva, J., Steinberg, L., & Belsky, J. (2007). Child-care history, classroom composition, and children's functioning in kindergarten. *Psychological Science, 18,* 1032–1039.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.

Epkins, C. C., & Meyers, A. W. (1994). Assessment of childhood depression, anxiety, and aggression: Convergent and discriminant validity of self-, parent-, teacher-, and peer-report measures. *Journal of Personality Assessment, 62,* 364–381.

Greenbaum, E., Decrick, R. F., Prange, M. E., & Friedman, R. M. (1994). Parent, teacher, and child ratings of problem behaviors of youngsters with serious emotional disturbances. *Psychological Assessment, 6,* 141–148.

Grimm, K. J. (2007). Multivariate longitudinal methods for studying developmental relationships between depression and academic achievement. *International Journal of Behavioral Development, 31,* 328–339.

Grimm, K. J., & Pianta, R. C. (2009). *A multitrait–multimethod analysis of the Child Behavior Checklist.* Unpublished manuscript.

Grimm, K. J., & Ram, N. (2009). Non-linear growth models in Mplus and SAS. *Structural Equation Modeling: A Multidisciplinary Journal, 16,* 676–701.

Hill, A. L., Degnan, K. A., Calkins, S. D., & Keane, S. P. (2006). Profiles of externalizing behavior problems for boys and girls across preschool: The role of emotion regulation and inattention. *Developmental Psychology, 42,* 913–928.

Hong, R. Y., Paunonen, S. V., & Slade, H. P. (2008). Big Five personality factors and the prediction of behavior: A multitrait–multimethod approach. *Personality and Individual Differences, 45,* 160–166.

Konold, T. K., & Pianta, R. C. (2007). The influence of informants on ratings of children's behavioral functioning. *Journal of Psychoeducational Assessment, 25,* 222–236.

Little, R. T. A., & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York, NY: Wiley.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing, 10,* 325–337.

Matson, J. L., & Nieminen, G. S. (1987). Validity of measures of conduct disorder, depression, and anxiety. *Journal of Clinical Child Psychology, 16,* 151–157.

McArdle, J. J., & Epstein, D. B. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58,* 110–133.

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods, 14,* 126–149.

McArdle, J. J., & Hamagami, F. (2001). Linear dynamic analyses of incomplete longitudinal data. In L. Collins & A. Sayer (Eds.), *Methods for the analysis of change* (pp. 137–176). Washington, DC: APA Press.

McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for early academic problems: The role of learning-related social skills. *Early Childhood Research Quarterly, 15,* 307–329.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika, 55,* 107–122.

Muthén, L. K., & Muthén, B. O. (2008). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

National Center for Education Statistics. (2001). *User's manual for the ECLS-K public-use data files and electronic codebook.* Washington, DC: U.S. Department of Education.

NICHD Early Child Care Research Network. (1997). The effects of infant child care on infant–mother attachment security: Results of the NICHD Study of Early Child Care. *Child Development, 68,* 860–879.

NICHD Early Child Care Research Network. (2000). Characteristics and quality of child-care for toddlers and preschoolers. *Applied Developmental Science, 4,* 116–135.

NICHD Early Child Care Research Network. (2001). Nonmaternal care and family factors in early development: An overview of the NICHD Study of Early Child Care. *Applied Developmental Psychology, 22,* 457–492.

NICHD Early Child Care Research Network. (2002). The relation of global first-grade classroom environment to structural classroom features, teacher, and student behaviors. *The Elementary School Journal, 102,* 367–387.

Ram, N., & Grimm, K. J. (2007). Using simple and complex growth models to articulate developmental change: Matching method to theory. *International Journal of Behavioral Development, 31,* 303–316.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Chicago, IL: University of Chicago Press.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203–228.

Stuhlman, M. W., Ferrer, E., Downer, J. T., Hamre, B. K., Mashburn, A. J., & Pianta, R. C. (2009). *Modeling the developmental dynamics between academic achievement and social behaviors across the elementary years.* Manuscript submitted for publication.

Wolfe, V. V., Finch, A. J., Saylor, C. F., Blount, R. L., Pallmeyer, T. P., & Carek, P. J. (1987). Negative affectivity in children: A multitrait–multimethod investigation. *Journal of Consulting and Clinical Psychology, 35,* 245–250.

Received February 27, 2009
Revision received August 27, 2009
Accepted November 25, 2009 ■