Heyman, G. D., & Giles, J. W. (2004). Valence effects in reasoning about evaluative traits. *Merrill-Palmer Quarterly, 50,* 84–109.

Hortacsu, N. (1987). Attributional focus and information selection in relation to chronological age. *Child Development, 58,* 225–233.

Kalish, C. (2002). Children's predictions of consistency in people's actions. *Cognition, 84,* 237–265.

Keil, F. C. (1989). *Concepts, kinds and cognitive development* (Learning, Development, and Conceptual Change series). Cambridge, MA: MIT Press.

Kelley, H. H. (1972). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 1–26). Morristown, NJ: General Learning Press.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28,* 107–128.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103,* 284–308.

Ladd, G. W., & Price, J. M. (1987). Predicting children's social and school adjustment following the transition from preschool to kindergarten. *Child Development, 58,* 1168–1189.

Livesley, W. J., & Bromley, D. B. (1973). *Person perception in childhood and adolescence.* London: Wiley.

McArthur, L. Z. (1976). The lesser influence of consensus than distinctiveness information on causal attributions: A test of the person–thing hypothesis. *Journal of Personality and Social Psychology, 33,* 733–742.

Menard, S. (2002). *Applied logistic regression analysis* (Quantitative Applications in the Social Sciences series, Vol. 106; 2nd ed.). Thousand Oaks, CA: Sage.

Mendelson, M. J., Aboud, F. E., & Lanthier, R. P. (1994). Personality predictors of friendship and popularity in kindergarten. *Journal of Applied Developmental Psychology, 15,* 413–435.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32,* 89–115.

Miller, P. H., & Weiss, M. G. (1981). Children's attention allocation, understanding of attention and performance on the incidental learning task. *Child Development, 52,* 1183–1190.

Nelson, S. A. (1980). Factors influencing young children's use of motives and outcomes as moral criteria. *Child Development, 51,* 823–829.

Newman, L. S. (1991). Why are traits inferred spontaneously? A developmental approach. *Social Cognition, 9,* 221–253.

Peevers, B. H., & Secord, P. F. (1973). Developmental changes in attribution of descriptive concepts to persons. *Journal of Personality and Social Psychology, 27,* 120–128.

Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development, 55,* 550–560.

Rholes, W. S., & Ruble, D. N. (1986). Children's impressions of other persons: The effects of temporal separation of behavioral information. *Child Development, 57,* 872–878.

Schuster, B., Ruble, D. N., & Weinert, F. E. (1998). Causal inferences and the positivity bias in children: The role of the covariation principle. *Child Development, 69,* 1577–1596.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Science, 7,* 246–251.

Smith, L. B. (2002). How to be smart: Lessons from word learning. In D. J. Lewkowicz & R. Lickliter (Eds.), *Conceptions of development* (pp. 259–278). New York: Psychology Press.

Stipek, D. J., & Daniels, D. H. (1990). Children's use of dispositional attributions in predicting the performance and behavior of classmates. *Journal of Applied Developmental Psychology, 11,* 13–28.

White, P. A. (1995). *The understanding of causation and the production of action: From infancy to adulthood.* Hove, England: Erlbaum.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology, 47,* 237–252.

Yuill, N. (1992). Children's conception of personality traits. *Human Development, 35,* 265–279.

Yuill, N. (1993). Understanding of personality and dispositions. In M. Bennett (Ed.), *The development of social cognition: The child as psychologist* (pp. 87–110). New York: Guilford Press.

Yuill, N., & Pearson, A. (1998). The development of bases for trait attribution: Children's understanding of traits as causal mechanisms based on desire. *Developmental Psychology, 34,* 574–586.

Received July 13, 2004
Revision received November 23, 2005
Accepted November 29, 2005 ■

Developmental Psychology
2006, Vol. 42, No. 3, 514–532

Copyright 2006 by the American Psychological Association
0012-1649/06/$12.00 DOI: 10.1037/0012-1649.42.3.514

Genetic and Environmental Influences on Academic Achievement Trajectories During Adolescence

Wendy Johnson, Matt McGue, and William G. Iacono
University of Minnesota—Twin Cities

Most studies have considered the effects of particular characteristics on academic achievement individually, which means that little is known about how they function together. Using the population-based Minnesota Twin Family Study, the authors investigated the effects of child academic engagement (interest, involvement, effort), IQ, depression, externalizing behavior, and family environmental risk on academic achievement (reported school grades) from ages 11 through 17. Hierarchical linear growth curve modeling showed main effects on initial reported Grades for all variables, and IQ mitigated the deleterious effects of family risk and externalizing. Only engagement affected change in Grades through adolescence. Influences on initial Grades were strongly genetically influenced, associated primarily with IQ, engagement, and externalizing behavior. Shared environmental influences on initial Grades linked engagement, IQ, and family risk. Genetic influences on change in Grades were substantial, but they were not associated with the academic, family risk, and mental health covarying factors. These results indicate that age 11 achievement and change in achievement through adolescence show systematic patterns and document the existence of individual differences in the commonly shared developmental experience of adapting to the school environment.

Keywords: genetic and environmental influences, school achievement, longitudinal twin study, achievement trajectories, covariates of achievement

Academic achievement is an important predictor of adult outcomes in our increasingly technological society. School achievement is associated with lower rates of negative outcomes such as teen pregnancy, welfare dependency, and criminal behavior, as well as higher levels of positive outcomes that include employment stability and lifetime income (Bronfenbrenner, McClelland, Wethington, Moen, & Ceci, 1996). To some degree, differences in adult outcomes no doubt reflect the greater opportunities associated with greater levels of academic attainment: for example, more and better-paying jobs are available to people with college diplomas than they are to those without college degrees. The process of educational attainment, however, is cumulative, with acquisition of basic skills in elementary school providing the foundation for knowledge acquisition in secondary school and good performance at that level key to school attendance beyond it. This reality points to the need to identify clearly the factors associated with academic achievement well before graduation from high school.

Factors Associated With Achievement

Prior studies have identified several classes of risk factors associated with achievement, including motivation (e.g., Eccles, Roeser, Wigfield, & Freedman-Doan, 1999), intelligence (e.g., Butler, Marsh, Sheppard, & Sheppard, 1985), psychopathology (e.g., Roeser, Eccles, & Sameroff, 1998, for depression; Hinshaw, 1992, for externalizing), and family environmental risk (e.g., Hedges & Nowell, 1999; White, 1982). These studies have generally considered the effects on achievement of these classes of characteristics individually, which means, however, that we know little about the extent to which they share covariance or transact with each other to influence achievement. One exception concerns the relationships among various kinds of externalizing behaviors and achievement. In this area, several studies (Fergusson & Horwood, 1995; Fergusson, Horwood, & Lynsky, 1993; Frick et al., 1991; Johnson, McGue, & Iacono, 2005) have shown that attention problems and low IQ can completely explain the well-established association between conduct and achievement problems. This relationship makes clear the need to investigate the possibility of similar kinds of interrelationships among other individual characteristics that, by themselves, are associated with achievement.

The contribution of family environmental risk to poor achievement illustrates this gap in our understanding in another way. Prior studies have identified a significant number of specific aspects of family environmental risk factors associated with achievement, including socioeconomic status (e.g., White, 1982), ethnicity (e.g., Hedges & Nowell, 1999), parental mental health problems (e.g., Sameroff, Seifer, Baldwin, & Baldwin, 1993), parental involvement and support (e.g., DuBois, Eitel, & Felner, 1994), and stressful life events such as experience of violence and parental break-up or job loss (e.g., Gutman, Sameroff, & Eccles, 2002). In general, however, recent research has tended to confirm Rutter's (1979) observation with respect to psychiatric disorder that no particular, single aspect of environmental risk predicts outcome. Rather, the

Wendy Johnson, Matt McGue, and William G. Iacono, Department of Psychology, University of Minnesota—Twin Cities.

This research was supported by U.S. Public Health Service Grants AA00175, DA13240, and DA05147. We thank the twins and their families and the recruiting, interviewing, data management, and laboratory staffs of the Minnesota Twin Family Study.

Correspondence concerning this article should be addressed to: Wendy Johnson, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455. E-mail: john4350@tc.umn.edu

presence of multiple areas of risk leads to the outcome because the breadth of risk increases the probability of a negative outcome (Wachs, 2000). This finding is at least partly true because individual characteristics and environmental risk factors tend to cluster in the same families in a way that reflects severity of risk (Masten & Coatsworth, 1998).

It seems likely that family risk factors tend to cluster as they do, in part, because they both have common causes and are mutually influential. For example, low parental education, poor parental mental health, and single parenthood likely each contribute directly to parental poverty and to poor parental involvement and support, thus adding to the burden of risk borne by the offspring of poor parents. These circumstances do not occur randomly; rather, they reflect characteristics of individual family members to some degree. Low parental intelligence contributes directly to low parental education (Ceci & Williams, 1997), and certain personality characteristics such as negative affect and lack of constraint contribute directly to poor parental mental health (Krueger, Caspi, & Moffitt, 2000; Krueger, Caspi, Moffitt, Silva, & McGee, 1996) and to single parenthood (Jockin, McGue, & Lykken, 1996). In addition, large volumes of data clearly indicate substantial genetic influences on intelligence (Bouchard & McGue, 1981), personality (McGue, 2001), and mental health (Krueger, 1999). These data suggest that, in addition to the environmental risks faced by children growing up with parents with these characteristics, these children also face genetic risk for these same characteristics themselves. If we hope to disentangle the causative roles played by the variables associated with achievement, we need to acknowledge the implications of these data by making use of samples in which we can address the question of underlying genetic and environmental relationships. Thus one of the primary purposes of this study is to use a population-level, genetically informative sample to investigate the nature of multivariate associations among a range of variables that have been linked individually to academic achievement.

Patterns in Achievement Over Time

A high degree of continuity from year to year exists in the relative achievement of individuals, but the correlations over time are far from complete. For example, Jimerson, Egeland, and Teo (1999) found correlations of approximately .6 in achievement test scores for the period from early to middle childhood and middle childhood to high school. Though substantial, correlations of this magnitude leave plenty of room for change, as well as continuity. The existence of such change, along with the increasing social policy emphasis on seeing "no child left behind," has led recent researchers to investigate academic achievement trajectories over time (Gutman, Sameroff, & Cole, 2003; Jimerson et al., 1999; Kowalski-Jones & Duncan, 1999; Marston & Tindal, 1995). The goal of such investigations has generally been to determine which factors are associated with change so as to understand why some children fall further and further behind their peers while others suffer temporary perturbations from which they recover. In general, the conclusions have been that change is common, but much of it acts idiosyncratically. One possible explanation for this pattern might be the presence of a large amount of effectively random and self-correcting change in the short term, but the relatively small amount of change that does persist operates in a lawful fashion related to variables that are already present in early childhood.

Gutman et al. (2003) made use of grade point averages (GPAs) to measure achievement from grades 1 through 12 in 145 participants in the Rochester Longitudinal Study. This method means that these authors, unlike researchers who make use of achievement test scores, did not expect to observe changes in average achievement from period to period, as a child growing in skills and knowledge at the expected rate would be expected to earn the same GPA from year to year. Nevertheless, the authors observed decreasing GPAs over time, as have other studies (Eccles, Roeser, Wigfield, & Freedman-Doan, 1999). Greater mental health and lower family risk (i.e., higher socioeconomic status and better parental mental health) at age 4 predicted slower rates of decrease. In addition, main effects of family risk, IQ, and mental health were factors on initial achievement, and both IQ and mental health interacted with family risk so that, across the ranges of the other covariates, when family risk was low, IQ and mental health were associated with higher achievement but not when family risk was high. Thus, in this study, individual characteristics offered protective effects only to those in relatively good environmental circumstances, providing little explanation for the resilience that some children in poor circumstances display (Garmezy, 1993). We had hoped to investigate this factor further in the current study.

The Gutman et al. (2003) study did show, however, that level of and change in achievement appeared to have patterns over time and that these patterns were linked with the same individual and family level characteristics that have been associated with achievement at particular points in time. This finding suggests that, despite the large amount of idiosyncratic change, some patterns are systematic (in the sense that they can be captured by linear or exponential equations, or both), thus documenting important individual differences in the commonly shared developmental process of adapting to the school environment. Illuminating the nature of the genetic and environmental influences on these patterns and associations should help develop understanding of the mechanisms involved in individual differences in the achievement outcome, which is another purpose of the current study.

Gender Differences in Achievement

Substantial mean gender differences exist in academic achievement, as well as in the individual characteristics that are associated with it. In general, girls receive higher grades than boys do and score more highly on subject matter achievement tests, from elementary school through college (e.g., Kimball, 1989; Mau & Lynn, 2001), in spite of the fact that boys tend to score higher than girls do on college and other aptitude tests (Mau & Lynn, 2001). One result of this discrepancy is that more girls than boys end up attending college at present in the United States (American Association of University Women, 1996). Effect sizes for the gender differences in achievement vary considerably, depending on the measure used and the age of the students, but can range as high as .5 standard deviation. Effect sizes for the gender differences in aptitude are generally very small and may be nonexistent for intelligence per se (Deary, Thorpe, Wilson, Starr, & Whalley, 2003). Greater involvement with the social environment of school through motivation or engagement may explain part of girls' advantage in achievement (Hyde & Kling, 2001), given that girls

tend to respond more to the externally assigned value of an achievement task than do boys (Eccles, 1984). Externalizing problems, including both disruptive behavior and inattention, may also explain part of the difference because these problems are much more common in boys than they are in girls (e.g., Butts et al., 1995; Cohen et al., 1993, for disruptive behavior; Gomez, Harvey, Quick, Sharer, & Harris, 1999; Rhee, Waldman, Hay, & Levy, 2001, for attention problems), with ratios of boys to girls on the order of 4:1 to 5:1 commonly observed. In addition, family risk factors tend to affect boys more severely than they do girls (e.g., Hetherington, 1989), which may contribute further to the difference. The presence of depression might serve to offset these effects to some degree, given that the prevalence of depression is generally higher in girls than it is in boys (e.g., Roberts & Vernon, 1983), at least after puberty. The interrelationships of these factors with achievement in girls and boys have not, however, been investigated to our knowledge. To do so was another purpose of the present investigation.

Overview of Current Study

In this study, we made use of the 11-year-old cohort of the Minnesota Twin Family Study, a population-based longitudinal study, to investigate three hypotheses suggested by the synopsis previously mentioned. First, we expected that several factors, including engagement, intelligence, family risk, externalizing behaviors, and depression that have been separately identified in prior research, would have independent main effects on level of achievement even when considered together. At the same time, we expected that these effects would be reduced from the main effects observed for the variables when measured alone, and we were interested in which variables would retain predictive importance in the multivariate context. We were also interested in possible interactive effects on achievement among these variables. We predicted that we would observe interactions involving IQ, mental health, and family risk but that the somewhat unexpected, specific interactive effects observed by Gutman et al. (2003) suggesting that individual characteristics only offer protective effects in lower risk environments would not replicate. Rather, we predicted that individual characteristics would offer protective effects in higher-risk environments. Second, we expected to replicate observations that achievement decreases over time, though we were agnostic regarding how this change might be related to the academic, engagement, family risk, and mental health factors we were considering because existing research in this area was limited. Third and finally, we expected that girls would have a higher initial level of achievement than would boys and that girls' achievement might decrease less with time.

In addition to testing these specific hypotheses, we addressed two exploratory research issues. First, we estimated the magnitudes of genetic and environmental influences on initial achievement and change, and we investigated the relationships among the genetic and environmental influences on the variables from a descriptive perspective. Second, we compared the resulting biometric models for girls and boys in an effort to identify differences in genetic and environmental influences or the relationships among them that might contribute to the differences in achievement that have been observed.

Methods

Sample

Participants were drawn from the 11-year-old cohort of the Minnesota Twin Family Study (MTFS), an ongoing longitudinal study of a community-based sample of like-sex twins and their parents. The MTFS sample was recruited using a population-based method. The current status and location of more than 90% of the like-sex twin pairs born in Minnesota in the targeted years was determined starting from birth records and using various publicly available databases. Located twins living within a day's drive of Minneapolis with at least one biological parent and who had no mental or physical handicap precluding participation were invited to complete a day-long, in-person assessment at our laboratories at the University of Minnesota; approximately 20% declined to do so. In addition, more than 80% of the families who did not participate completed a brief mail or telephone survey, which enabled some comparison of participants and nonparticipants. This comparison revealed that parents in participating families were significantly, though only modestly, better educated than those in nonparticipating families, with a mean difference of less than 0.3 years of education. The two groups of families did not differ significantly in self-reported mental health. The MTFS intake sample is thus generally representative of families with twins born in Minnesota. A complete description of the ascertainment and assessment procedures used in the MTFS, as well as an analysis of nonparticipants, is given in Iacono, Carlson, Taylor, Elkins, and McGue (1999). We based our assessment of the twins' zygosity on the consensus of (1) parental reports of physical similarity (e.g., eye and hair color), (2) an algorithm using twin similarity of ponderal and cephalic indices and fingerprint ridge, and (3) the judgment of two senior staff. When disagreement occurred, serological analysis was carried out. We evaluated the reliability of the consensus procedure by comparing serological and consensus results for 50 pairs of twins. The two sets of results showed complete agreement.

The sample is in no way enriched for families at high risk; neither is it limited to families facing little risk. Rather, the sample generally reflects the full range of risk-associated characteristics present in the native Minnesota population as a whole. The average Hollingshead occupational level for the families was a little less than 4, indicating possession of jobs that required some education just beyond the skilled "blue collar" level. The sample included parents working in highly professional occupations, as well as parents who were unemployed or working in semiskilled jobs (the standard deviation was just under 2 Hollingshead levels; range 1 to 8, with 1 indicating professional, 7 indicating unskilled, and 8 indicating unemployed). Our estimates of genetic and shared and nonshared environmental influence can be considered representative of those in populations of this type.

The intake assessment of the 11-year-old cohort was administered when the twins were, on average, the age of 11 years, though a few were not quite 11, and some had recently turned 12 at time of assessment. Data were available for 443 pairs of girls (269 monozygotic [MZ], 173 dizygotic [DZ], 1 missing zygosity) born from 1981 to 1984 and from 1988 to 1990 and 381 pairs of boys (256 MZ, 122 DZ, 3 missing zygosity) born from 1977 to 1982. Thus a small cohort difference (3 to 5 years) exists that might appear to confound our observation of gender differences. Some overlap occurs in the birth years for girls and boys, however (1981 and 1982), and no difference was noted in reported grades for the girls and boys in versus out of the overlap period, suggesting that the cohort difference does not confound gender comparisons. Consistent with the demographics of Minnesota for the birth years sampled, over 95% of the twins were Caucasian. The first full follow-up assessment was made when the twins were 14 years of age, though, again, some were just short of 14 at time of assessment, and some had recently turned 15. Data were available for 712 individual girls (80.4%) and 707 individual boys (92.8%). The proportion of girls with follow-up data was so much lower than that for boys because 124 individual girls were born during 1988 to 1990, and their follow-up assess-

ments had not been processed at the time of this analysis. Nevertheless, data from these girls are informative for the baseline correlates of academic achievement and might be efficiently analyzed using full information maximum likelihood techniques. Without these younger girls, the percentage of girls participating at the first follow-up was 93.4%. The second full follow-up assessment was made when the twins were approximately 17 years of age. At this time, data were available for 507 individual girls (57.2%) and 625 individual boys (82.0%). Again, the proportion of girls with data was much lower than that of boys because many of the girls had not yet reached age 17.

In addition to the full follow-up assessments, very abbreviated assessments were administered to parents (by telephone) and teachers at approximately annual intervals between the full assessments. These abbreviated assessments included parent and teacher reports of child grades (see later discussion). Because of cuts in funding that resulted in the termination of these abbreviated assessments, participation rates for the annual assessments were substantially lower than they were for the full assessments, particularly for the abbreviated assessments between the first and second full follow-ups. Participants in the first follow-up and second follow-up did not differ on reported grades from those who dropped out after intake, though reported grades from the abbreviated assessments were higher than those from the intake and full follow-up assessments. Twins with annual data reports also tended to have lower rates of depression and lower family risk than did those without, though they tended to display higher rates of disruptive behavior. The extent to which this tendency was true varied, however, from report to report, and expectation maximization analysis of the full grade covariance matrix indicated that the hypothesis that the data were missing completely at random could not be rejected (Little's Missing Completely at Random Test, $\chi^2 = 108.64$, $N = 1639$, 115 df, $p = .65$), indicating that the differences between those with and without annual data reports did not present analytical difficulties. Although the amount of grades data available at ages 15 and 16 was minimal, such situations can be well accommodated by full information maximum likelihood, and the result is greater precision in the estimates of change in grades (Little & Rubin, 1987). We thus made use of the full raw grade data throughout the analyses we present. We also prepared the analyses omitting ages 15 and 16 from the models completely, with highly similar patterns of results. The numbers of participants providing data at each annual report are shown in Table 1.

Measures

Reported grades. MTFS collects data across many domains from as many as three reporters: twins, parents, and teachers. Twin and parent reports are provided during structured in-person interviews in our laboratories, during structured telephone interviews, and by self-report questionnaires completed in our laboratories and at home. Teacher reports are obtained by having twins nominate as many as four different teachers and asking these teachers to complete an extensive questionnaire of student behavior and achievement. Minnesota state guidelines stipulate that twins should be in different classrooms whenever possible; thus co-twins usually do not nominate the same set of teachers. Teachers represent an important source of information because they see the twins in a normative, structured setting on a frequent basis away from their immediate families and have no a priori emotional ties to them. The teachers are thus able to put the twins' behavior in a broader and less biased context than are parents. In addition, the teachers assign some of the grades on which they report. At the same time, however, the twins' behavior may be somewhat more constrained at school than it was at home, given that the setting at school is public. Return rates on the Teacher's Rating Forms exceeded 70%.

The study does not, in general, collect data on actual grades because of the disparities in grading formats, procedures, and standards in the various school systems from which the MTFS families are drawn. Rather, parents, twins, and teachers report separately on student grades in language arts, math, social studies, and science classes, as well as overall by indicating that the grades are much better than average (A = 4), better than average (B = 3), average (C = 2), below average (D = 1), or much below average (i.e., failing = 0). For teachers' reports, the estimated internal consistency reliability for the grade reports was .92, and estimated interteacher agreement reliability was .87. Overall, the correlations among the grade reports for the three categories of reporters and the three full assessment time points were high, reflecting both the stability over time in reported grades and reliability among reporters. Correlations among reporters within time period ranged from .57 to .83, with an average correlation of .73. Correlations across time for the same rater were somewhat lower; they ranged from .41 to .70, with an average of .59. Thus our measurements were apparently sensitive enough to distinguish change over time from rating inconsistency. For this study, we made use of a composite across available reporters of the average of the grades in each subject,[1] with possible scores thus ranging from 0 to 4. As a check on the validity of our composite reporter measure, we used school transcripts provided for a random sample of 67 participants assessed at age 14. The correlation between GPA computed from school transcripts and our composite reporter measure was .89, indicating excellent correspondence between measures.

To justify the appropriateness of aggregating the data over raters to construct the composite grades variable, we constructed a measurement model of the individual average grades by reporter for each of the full assessments and evaluated this model separately for girls and boys. The model consisted of a general factor representing average grades and factors specific to each time period. The addition of factors specific to each reporter did not generate factors with substantial loadings, nor were significant correlations by reporter found in the residual variances. We were able to obtain strict factorial invariance (Meredith, 1993) across gender in this model, justifying its use in comparing results between girls and boys. The presence of strict factorial invariance provided another validity check that the composite we computed would measure grades adequately when compared with a latent grades variable. Use of the composite streamlined our analysis considerably. For convenience, we will refer to this composite report measure as *Grades*.

Engagement at age 11. Twin school engagement was assessed using questions from a self-report questionnaire on school behaviors in which participants reported the degree to which they were psychologically involved with school. Items included interest in schoolwork, study without being reminded, homework completion, enjoyment of attending school, and desire for good grades, rated on a 4-point scale ranging from "1. definitely true of me" to "4. definitely false of me." We scored these items so that high scores reflected high engagement and summed them. Possible scores ranged from 4 to 16. Estimated internal consistency reliabilities for the scale were .74 for girls and .85 for boys.

IQ. The twins were assessed using an abbreviated version of the Wechsler Intelligence Scale for Children—Revised (WISC–R; Wechsler, 1974) at age 11 consisting of two verbal (Vocabulary and Information) and two performance (Block Design and Picture Arrangement) subtests. These subtests were selected for their high correlation (.90) with total WISC–R IQ based on all subtests.

[1] In computing this composite, we first averaged the subject grades reported by each teacher for each participant and then averaged the multiple teacher reports (using the actual number of teacher reports received) to obtain an average teacher report. We then separately averaged the grades reported by parents and children by subject and then averaged across reporters to generate a straightforward continuous measure most directly analogous to GPA.

Table 1
Descriptive Statistics and Effect Sizes of Mean Differences and Their Significance for Girls and Boys

	Girls			Boys			Gender differences	
Measure	*M*	*SD*	*N*	*M*	*SD*	*N*	Effect size	*p* value
Average grades								
Age 11	3.16	0.65	880	2.85	0.70	759	−.45	<.001
Age 12	3.17	0.79	433	2.90	0.83	740	−.33	<.001
Age 13	3.24	0.74	267	2.85	0.87	594	−.46	<.001
Age 14	3.13	0.74	712	2.79	0.84	707	−.42	<.001
Age 15	3.15	0.84	61	3.07	0.91	111	−.09	.560
Age 16	3.15	0.86	29	3.17	0.93	60	.02	.922
Age 17	3.21	0.69	507	2.82	0.83	625	−.49	<.001
Age 11 covarying factors								
Engagement	13.91	1.41	834	12.87	1.75	754	−.28	<.001
IQ	102.44	13.97	833	104.62	13.65	750	.16	.002
Family risk	3.07	2.44	828	2.73	2.29	692	−.08	.006
Externalizing	2.81	3.22	822	4.73	4.78	754	.81	<.001
Depression	1.79	4.73	868	.55	1.77	754	−.74	<.001

Note. Effect size is the mean difference divided by pooled standard deviation, stated so that boys higher is positive. Engagement was measured using a scale that ranged from 4 to 16, as reported by the children. IQ was measured using prorated scores from four scales of the WISC-R (Wechsler, 1974). Family risk was a cumulative measure reflecting parenting, parental mental health, and family demographics. Externalizing and depression were measured using structured interview-based symptom counts. See text for further description of measures. *M* = mean; *SD* = standard deviation; *N* = number of participants.

Family risk. We clustered 22 variables to produce a single measure, broadly tapping many elements of family risk.[2] Table 2 shows the variables we selected. Use of such a composite measure makes the consideration of a large number of predictor variables in a single analysis possible, and it provides a good summary of effects without sacrificing degrees of freedom when a single outcome is of interest (Ackerman, Izard, Schoff, Youngstrom, & Kogos, 1999), especially in longitudinal analyses (Deater-Deckard, Dodge, Bates, and Pettit, 1998). Any cumulative measure of this type involves many judgmental decisions about the levels of the variables that will be used to signify family risk. For some of the variables we selected, such as parent race, single-parent status, and parent level of education, thinking in terms of a single practical threshold is relatively reasonable. Thus, for our measure, we tabulated presence in the home of a biological or stepmother or stepfather of any race other than Caucasian (5%); family size larger than five children (7.5%); twins living with a single parent (either mother or father = 9%); number of parents with less than a high school education (two possible; 2% had at least one); either father with a Hollingshead occupational code (Hollingshead, 1957) of at least 6 (semiskilled) or mother with a Hollingshead occupational code of 7, or both (approximately 3% of the sample).

For other variables, however, the judgments about level of family risk are much more arbitrary, particularly when the variables can be considered continuous. Table 1 describes the judgments we made for each of the variables we considered. To generate our measure of family risk, we summed the presence of all 22 risk variables. Possible scores thus ranged from 0 to 22. Though many of the risks summed were the same for members of a twin pair, some, such as birth weight, parenting quality, and life events, varied for individual twins. Thus each twin had a separate score for family risk. To examine the effects on the overall family risk variable of the judgments we made in constructing it, we constructed four alternative family risk variables using different percentages of the distributions of the continuous variables and other criteria for determining the thresholds for these variables. Even when different criteria were used for several different variables at the same time, the correlations of these alternate family risk variables with the family risk variable we used remained above .90 for both girls and boys.

Externalizing. Twins and parents completed the Diagnostic Interview for Children and Adolescents (DICA; Welner, Reich, Herjanic, Jung, & Amado, 1987) to assess attention deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder present in each twin at age 11. If either the mother or the child reported a symptom, it was assigned a value of 2 for a rating of "definitely present," and 1 for a rating of "possibly present," and the values were summed across the three disorders to yield "externalizing" scores, possible scores thus ranging from 0 to 56.

Depression. An analogous procedure was followed to obtain symptom counts of major depressive disorder, again assigning a value of 2 for a rating of "definitely present," and 1 for a rating of "possibly present." Possible scores ranged from 0 to 18.

[2] Clustering of risk factors within families as described previously introduces measurement difficulties in two ways. First, colinearity among measures of risk is inevitable, and second, the clustering makes isolating the unique effects of any specific family risk factor difficult and unimportant, especially in relatively small samples. The clustering does, however, suggest the pragmatic analytical strategy we used. That is, we developed dichotomized scores denoting presence (1) or absence (0) of family risk based on the continuous scores on each of the 22 risk conditions and summed them. This strategy wastes some information on variability in family risk indices (Burchinal, Roberts, Hooper, & Zeisel, 2000) and weights each family risk factor evenly, which may or may not be accurate. It has the advantage, though, of making it possible to consider a large number of predictor variables comparing multiple regression and cumulating strategies directly. Deater-Deckard, Dodge, Bates, and Pettit (1998) compared multiple regression and cumulating strategies directly. They found that the cumulating strategy explained approximately two-thirds as much variance as multiple regression in cross-sectional analyses but did better in longitudinal analyses, suggesting that the advantage of regression-based methods may be capitalization on sample-specific features of the data that do not generalize over time. Moreover, the results of the cumulating strategy were not very sensitive to the specific family risk factors included. This strategy has been used successfully in recent studies related to achievement (e.g., Gutman, Sameroff, and Eccles, 2002).

Table 2
Percentages of Girls and Boys Above and Below Thresholds Set for the Variables Contributing to the Family Risk Composite

Variable	% of girls above threshold	% of boys above threshold	% overall above threshold
Relatively easily dichotomized variables			
Race	5.2	4.5	4.9
Large family	7.3	7.7	7.5
Single parent	8.7	9.5	9.1
Low education: mother	.9	1.4	1.1
Low education: father	1.7	1.1	1.4
Low job status: mother	2.5	2.5	2.5
Low job status: father	13.4	7.0	10.4
Major variables less easily dichotomized			
Low parental involvement: mother	17.6	18.7	18.1
Low parent regard for child: mother	17.6	20.3	18.9
Low parental structure: mother	21.7	20.2	21.0
Low parental involvement: father	20.1	17.8	19.0
Low parent regard for child: father	18.3	19.6	18.9
Low parental structure: father	17.9	16.3	17.2
Many stressful life events	21.3	12.4	17.2
Adult antisocial behavior: mother	19.3	20.7	19.9
Major depressive disorder: mother	23.5	21.6	22.6
Drug abuse: mother	19.7	19.2	19.5
Adult antisocial behavior: father	13.9	22.7	18.0
Major depressive disorder: father	17.9	20.3	19.0
Drug abuse: Dad	18.2	18.9	18.5
Minor variables less easily dichotomized			
Low birthweight	9.2	10.7	9.9
Interviewer assessment of parent	0.9	1.1	1.0

Note. Parental structure, involvement, and regard for child were composites of parents' and child's reports from the Parental Environmental Questionnaire (Elkins, McGue, & Iacono, 1997). To obtain information about stressful life events, we extracted questions regarding the occurrence of events that we judged most children would find stressful from the in-person "Structured Life Events Interview" completed by the twins. The events we extracted included moving to a new school district, having close friends move away, deaths of close relatives or friends, parents experiencing periods of intense conflict or separation, and new adults coming to live with the family. Parental mental health was assessed using the "Structured Clinical Interview" from the *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed., (*DSM-IV*; American Psychiatric Association, 1994). We made use of symptom counts for "adult antisocial behavior" (i.e., *DSM* antisocial personality disorder criteria reflecting antisocial behavior occurring since age 15), "major depressive disorder," and "drug abuse" or dependence of any form. For these major variables that are not easily dichotomized, we assigned presence of risk to approximately the top 20% of each variable's distribution. We judged birthweight to be a risk factor when it was less than 64 ounces; twins commonly have relatively low birthweights. Minnesota Twin Family Study (MTFS; Iacono et al., 1999) interviewers rated parents on 14 items reflecting parenting quality at the end of our assessment day. We assigned presence of family risk to those scoring in the bottom 30% on all items. We judged these variables to have relatively less impact than the others except in the more extreme ranges of their distributions.

Analytical Approach

Phenotypic analyses to address the first three hypotheses. To reduce positive skewness, we log-transformed the externalizing and depression symptom counts (after adding 1) before standardizing all variables so that they were centered at 0, with standard deviations of 1. The other variables were reasonably symmetrically distributed so that transformation was not warranted (e.g., the range of reported Grades at age 11 was 1 to 4, with skewness of −.55). We made use of maximum likelihood estimation, as implemented in *LISREL 8.53* (Jöreskog & Sörbom, 2002) to fit a hierarchical linear model (HLM) designed to address our research questions relating to phenotypic relationships among initial Grades, change in Grades over time, and covariate factors. Because we were working with twin pairs, we used a three-level HLM. The units of analysis at the first level were reported Grades over time. The units at the second level were the academic, family risk, and mental health covarying factors as they applied to twin pairs. We treated gender as another covarying factor here. Thus we compared phenotypic results for girls and boys by evaluating the significance of gender and gender × time interaction in the HLM regression equation. We also considered all the possible interactions between gender and each of the academic, family risk, and mental health covarying factors. The units at the third level were the individual twins within the twin pairs. Coefficients in the model reflect differences in Grades per standard deviation unit of change in the associated variable.[3]

[3] The regression equation for the first level of the HLM could be expressed as $Grades_{ti}=\beta_{0i} + \beta_{1i} + \varepsilon_{ti}$.

The second level was $\beta_{0i} = \alpha_{00} + \sum_{k=1}^{6}\alpha_{0k}x_{ki} + \sum_{h>j}^{6}\sum_{j=1}^{6}\gamma_{jk}x_{ji}x_{ki} + \mu_{0i}$, and $\beta_{1i}=\alpha_{1i}+\sum_{k=1}^{6}\alpha_{1k}x_{ki}+\mu_{1i}$.

For two covariates (IQ and family risk, for example), these would be expressed as $\beta_{0i} = \alpha_{00} + \alpha_{01}IQ_i + \alpha_{02}Risk_i + \gamma_{12}IQ_iRisk_i + \varepsilon_{ti}$ and $\beta_{1i} = \alpha_{1i} + \alpha_{11}IQ_i + \alpha_2Risk_i + \mu_{1i}$.

Analyses of genetic and environmental influences to address the last two research issues. The standard univariate quantitative genetic model is based on the assumption that the observed phenotypic variance (Vp) is a linear additive function of (independent) genetic (A) and shared (C) and nonshared (E) environmental components of variance. Symbolically,

$$Vp = A + C + E.$$

Under this model, the shared environmental variance represents experiential factors common to the members of a twin pair and operating to make them similar. They include experiences such as growing up with the same depressed or drug-abusing parents and parental socioeconomic status. For instances when MZ twins elicit more similar experiences than did DZ twins from their environments because of their greater genetic similarity, however, this is generally considered to be an expression of their genetically influenced characteristics. Nonshared environmental influences are specific experiential factors unique to each member of a twin pair and operating to make them different. Such experiences may include injuries and illnesses, having different teachers, participating in different leisure activities such as sports, and receiving different parental treatment. The distinction between the two is subtle. For example, two children in the same family may experience the same event (e.g., parental divorce), but this event is only a shared environmental influence to the extent that it makes the children similar—they may react to it very differently. The nonshared environmental component also includes variance attributable to measurement error. Genetic variance can be additive in the sense that, if multiple genes influence the trait, they do so independently of each other. This variance can also be nonadditive, reflecting dominance and other interactive polygenic effects. Given only additive genetic effects, the expected covariance between any two members of a twin pair as a function of the variance components given previously can be specified as,

$$COV_{(MZ)} = A + C$$

$$COV_{(DZ)} = .5 \times A + C.$$

We did not fit models that included nonadditive genetic effects because no evidence existed that they were relevant as the DZ correlations were more than one half the MZ correlations for all the variables.

The standard univariate model can be extended to multivariate situations by modeling the covariance between one twin's score on one variable and the other twin's score on another variable in a manner directly analogous to the univariate case. We began by making use of a quantitative genetic latent growth curve model fit with the structural equation modeling program, called *Mx* (Neale, Boker, Xie, & Maes, 1999). This model decomposes the intercept variance, slope variance, and the intercept-slope covariance into genetic and shared and nonshared environmental components, providing baseline estimates of overall genetic and environmental influences on initial Grades (intercept) and linear change in Grades (slope). We fit the model separately to the data for girls and boys, estimating genetic and environmental influences on the unexplained residual variances as well. Figure 1 diagrams the initial model. To estimate the extent to which genetic and environmental influences on the academic, family risk, and mental health covarying factors at age 11 might account for the genetic and environmental influences on initial Grades and linear change in Grades over time, we developed the model shown in Figure 2. This model is effectively equivalent to the HLM model used to describe the phenotypic Grades, except that the genetic and environmental portions of the latent phenotypic variance are decomposed, which makes estimating the proportions of the phenotypic correlations among the variables that can be attributed to genetic and environmental influences possible. Because depression at age 11 was not a significant phenotypic predictor of Grades in the HLM model, we dropped it from the model in Figure 2. We compared results for girls and boys by comparing model fit when their latent variable paths were constrained equal.

Results

Descriptive Statistics

Table 2 shows the means, standard deviations, and effect sizes of gender differences for each of the measures we used, separately for girls and boys. As expected, girls earned higher grades than did boys, and the effect size for most of the reporting period was moderate. In keeping with prior research, girls also showed considerably more depression symptomatology and somewhat higher academic engagement; they also experienced very slightly more family risk. Boys had slightly higher IQs and showed much more externalizing symptomatology. For these variables, no significant mean differences were found between MZ and DZ twins for either girls or boys.

Table 3 shows the correlations among the variables we used, separately for girls and boys. The correlations among reported grades over time were rather high, reflecting a high degree of stability. Little evidence was found for any difference in relative stability between girls and boys. Grades were moderately related to both engagement and IQ. The correlations between family risk and Grades and externalizing and Grades were somewhat lower and varied somewhat more over time. Depression was not significantly correlated with Grades in either girls or boys, nor did it have important relationships with any of the other academic, family risk, and mental

---

The third level expressed the individual twin variation around the pairvalues. The first-level model indicated that an individual's (indexed by $i = 1, \ldots, N$) Grades at any specific time (t) were a function of an initial level effect (β_{0i}), a linear change effect (β_{ti}), and a residual (ε_{ti}). At the second level, individual differences in the intercepts and slopes were modeled in terms of main effects of the six (x_{ki}; $k = 1, \ldots, 6$) covarying factors including gender (α_{0k}'s and α_{1k}'s respectively), and the interactions among the covarying factors including gender (γ_{jk}'s). We did not model interaction effects on the slopes.

In these equations, i referred to the twin pair; t to time in years; t was coded from 0 to 6, thus reflecting wave of data collection and making the β_0 (intercept) term interpretable as the grand mean for boys at age 11 when the covarying factors are at their mean levels. The fixed effects parameters of the model were the β's, and the α's were the effects of the academic, family risk, and mental health covarying factors on initial Grades and the effects of these factors on linear change in Grades. The random effects parameters were the μ_{ik} terms. The double summation in the second-level equation referred to the possible two-way interactions. We estimated the equation using both forward and backward elimination techniques, beginning the forward techniques with each covariate variable separately. We assessed model fit using the significance of the change in $-2 \times$ log-likelihood rather than the significance of the t statistics of the regression coefficients to increase the stability of the resulting model. The purpose of using both forward and backward techniques was to minimize the possibility that we were capitalizing on chance, as we required each variable to be significant regardless of which other variables were in the model, with the exception that we always retained main effects when assessing the significance of interactions. We tested all possible two-way interactions because we had little theoretical rationale for expecting particular interactions. We did consider the possibility of second-order main effects terms, but we found that none was significant and thus do not show these terms in the regression equations above. We checked HLM assumptions, including homoscedasticity and normality of residuals. The resulting model could be considered the most parsimonious representation of the effects we considered that had significant influences on Grades.

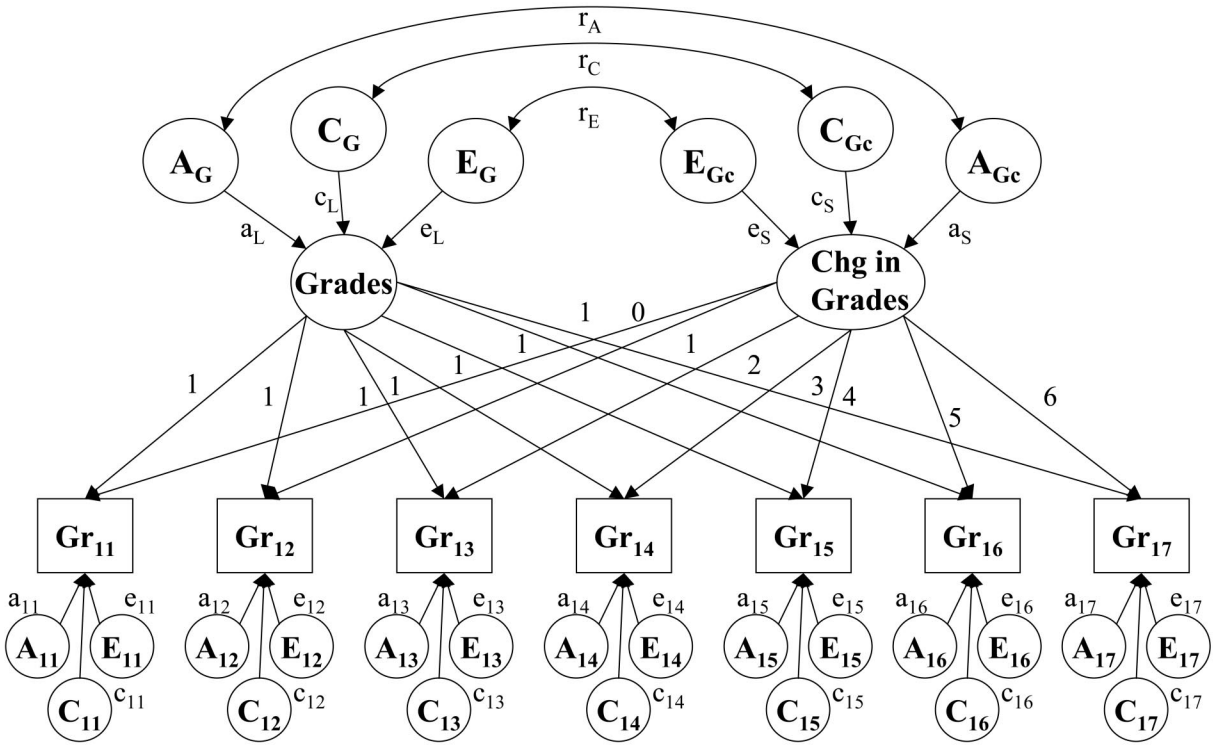

Figure 1. Basic gene-environment model for level of and change in achievement over time. A refers to genetic influence; C refers to shared environmental influence; E refers to nonshared environmental influence; subscripted numbers refer to age of measurement; lower case letters refer to estimated path coefficients and correlations; Gr_x refers to Grades at age *x*.

health covarying factors. We used the HLM model to develop a systematic way of interpreting these correlations.

Phenotypic Analyses

Figure 3 graphs the achievement trajectories for a random sample of approximately 5% of the data. The graph shows both the large individual differences in average Grades and the relative stability of Grades over time. At the same time, individuals with large fluctuations are also apparent.

Factors associated with initial level of Grades. To test our first hypothesis that several of our academic, family risk, and mental health covarying factors would have independent main effects on Grades when considered together, we fit the HLM model described previously. The most parsimonious model had the terms and coefficients reported in Table 4 stated such that the coefficients for the continuous covariates express the effects on Grades of a 1 standard deviation increase in the associated variable from the average. (Gender was stated with girls coded as 1, boys as 0.)[4] Thus the independent variables are standardized, but the dependent variable remains on a typical 0 to 4 GPA scale. Because of the potential for multicolinearity among the academic, family risk, and mental health covarying factors to influence the magnitude of the regression coefficients associated with them in the model, we show the coefficients for the main effects of the covarying factors two ways: (1) when they were estimated alone and (2) in the presence of all other main effects. This method allows comparison of the relative magnitude of the apparent effects when the covarying factors are considered in isolation and when they are considered together. All of the covarying factors appeared to be more important when considered alone than when considered together, indicating that they shared substantial variance. The variance shared was not sufficient to create statistical

[4] In addition to the overall sample coefficients shown in Table 4, HLM includes random effects components that reflect unexplained individual differences in the corresponding polynomial coefficients for initial Grades and change in Grades. For this model, individual differences in change that were not explained by the model could be captured by quadratic change terms, though the quadratic change term was not significant in explaining overall change in Grades. The random effects for a model with no academic, family risk, or mental health covarying factors were .3452 for initial Grades, .0174 for linear change in Grades, and .00041 for quadratic change in Grades. After including the academic, family risk, and mental health covarying factors, the random effects were .0955 for initial Grades, .0186 for linear change in Grades, and .00043 for quadratic change in Grades. Thus the academic, family risk, and mental health covarying factors served to explain 72% (1–.0955/.3452) of the variance in initial Grades. Adjustment for the effects of these covarying factors, however, actually acted to increase the relatively smaller variance at the individual level in linear and quadratic change in Grades. Part of the explanation for this is probably that the covariate factors measured at age 11, particularly engagement and externalizing behavior, are not likely to be very stable in adolescents in this age group.

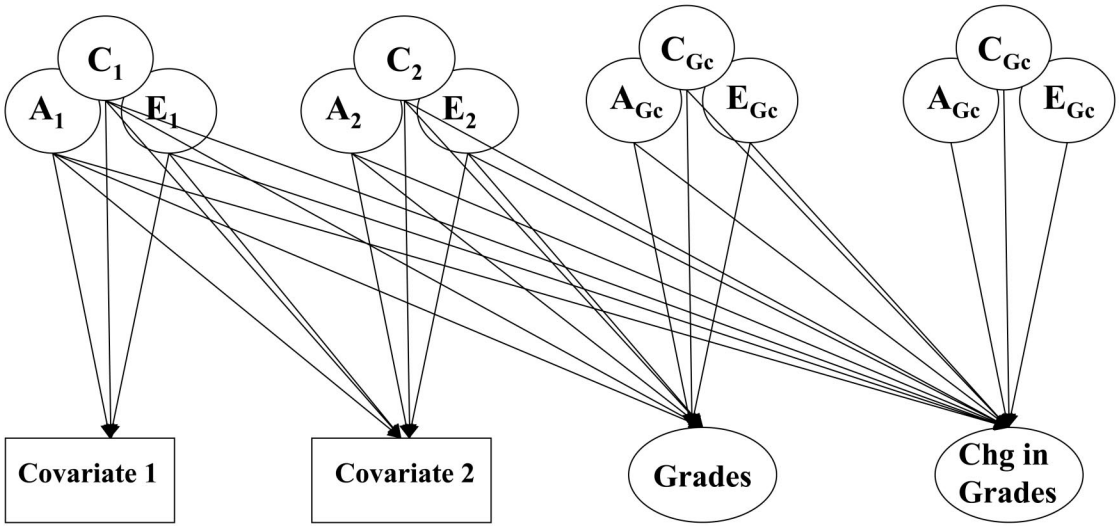

Figure 2. Extended gene-environment model for Grades, change in Grades, and covarying factors. A refers to genetic influence; C refers to shared environmental influence; E refers to nonshared environmental influence. The latent Grades variables are linked to the observed Grades variables as depicted in Figure 1. Only two covarying factors are shown here for simplicity. See Figure 1 for detail linking latent to observed Grades variables.

issues involving multicolinearity (the largest variance inflation factor was 1.35), but it was sufficient to affect the magnitude of the regression coefficients. The relative importance of gender, family risk, and externalizing was particularly diminished by the presence of the other variables in the model, suggesting that individual differences in these other variables (engagement and IQ [and to some degree externalizing]) largely account for the main univariate effects of gender, family risk, and externalizing.

Nonetheless, the model had the hypothesized independent main effects on initial Grades for gender and each of the academic, family risk, and mental health covarying factors, except for depression, which had no significant effect. Female gender, engagement, and IQ acted to increase Grades, and family risk and externalizing acted to decrease them. The main effect of female gender provided evidence in support of our hypothesis regarding gender differences. Only two significant interactive effects on initial Grades were noted: (1) between IQ and family risk and (2) IQ and externalizing. Each interaction was significant at $p < .05$ when in the model without the other, though the probability levels associated with both rose slightly above that when they were both included, because of multicolinearity between them. Because they were significant alone, we retained both interactions in the model. The three-way interaction was not significant, and no interactions involving gender was noted.

The interactions of IQ and family risk in predicting initial Grades meant that higher IQ mitigated the deleterious effects of family risk and externalizing. Thus, as we had predicted, we found interactions involving the same variables as did Gutman et al.

Table 3
Correlations among Study Variables

Measure	1	2	3	4	5	6	7	8	9	10	11	12
Average grades												
1. Age 11		**.75**	**.68**	**.67**	**.47**	**.54**	**.62**	**.52**	**.55**	**−.29**	**−.38**	**−.09**
2. Age 12	**.76**		**.78**	**.76**	**.42**	**.59**	**.65**	**.54**	**.52**	**−.25**	**−.41**	−.08
3. Age 13	**.69**	**.72**		**.77**	**.57**	**.60**	**.69**	**.53**	**.48**	**−.31**	**−.33**	**−.13**
4. Age 14	**.69**	**.70**	**.76**		**.72**	**.55**	**.70**	**.46**	**.54**	**−.34**	**−.34**	**−.12**
5. Age 15	**.60**	**.57**	**.68**	**.67**		**.82**	**.72**	**.12**	**.63**	**−.27**	**−.21**	−.01
6. Age 16	**.71**	**.58**	**.73**	**.68**	**.78**		**.73**	**.43**	**.56**	**−.16**	**.06**	**−.10**
7. Age 17	**.56**	**.58**	**.59**	**.71**	**.58**	**.81**		**.35**	**.48**	**−.24**	**−.26**	−.05
Age 11 covarying factors												
8. Engagement	**.56**	**.53**	**.55**	**.50**	**.29**	**.48**	**.39**		**.27**	**−.27**	**−.36**	−.08
9. IQ	**.50**	**.43**	**.39**	**.42**	**.48**	**.59**	**.40**	**.18**		**−.17**	**−.23**	**−.06**
10. Family risk	**−.22**	**−.21**	**−.24**	**−.20**	**−.11**	.07	**−.22**	**−.23**	−.06		**.28**	.15
11. Externalizing	**−.32**	**−.35**	**−.40**	**−.37**	**−.17**	.03	**−.30**	**−.43**	**−.10**	**.24**		.21
12. Depression	−.05	−.03	.01	−.04	−.02	**.12**	.00	−.06	.07	**.10**	**.17**	

Note. Correlations shown in bold were significant at $p < .01$. Girls are above the diagonal, boys below. Externalizing and depression were log-transformed to reduce skewness.

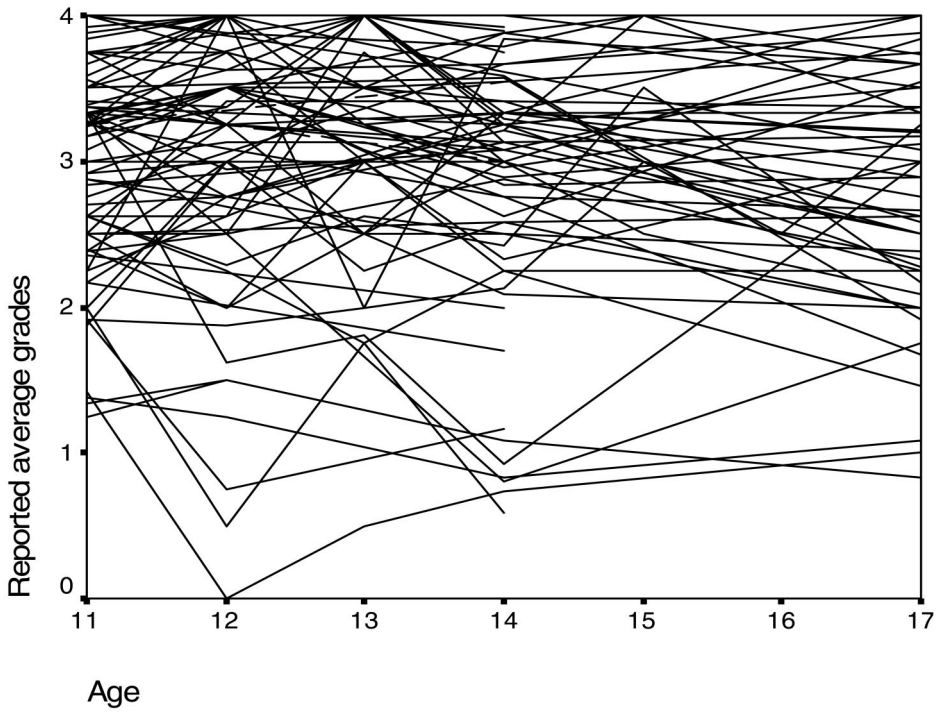

Figure 3. Random sample of approximately 5% of reported average Grade trajectories.

(2003), but their precise interactive effects did not replicate. This finding is illustrated in Figure 4, based on the coefficients from the regression equation. Our sample showed reasonable numbers with both high risk and high IQ and both low risk and low IQ (14 had both IQ [> 121] and risk in the top 10%, and 21 had both IQ [< 88] and risk in the bottom 10%). The interaction of IQ and externalizing had a similar impact: higher IQ acted to offset the effects of high externalizing, and low externalizing acted to offset the effects of low IQ; all else held equal at their means. This finding is also shown in Figure 4. Again, the sample included reasonable numbers with both high externalizing and high IQ (12 had both IQ and externalizing in the top 10%, and 23 had both IQ and externalizing in the bottom 10%). The effects of the interactions, however, were rather small.

Factors associated with change in Grades. We used the same HLM model to test our second hypothesis involving linear change in Grades. Again as hypothesized, Grades fell over time (0.172 Grade points per year in Table 4) for boys, after controlling for the relationships among the academic, family risk, and mental health covarying factors at age 11.[5] As shown in Table 4, female gender acted to offset the rate of decline (the 0.192 coefficient applicable to girls offset the −.172 decline experienced by all, producing a small net increase for girls), providing evidence for our third hypothesis that this result would be the case. Engagement, however, acted to accelerate the rate of decline (−0.083). The result for engagement was counterintuitive because it suggested that children with high engagement at age 11 saw greater decrease in reported Grades over time than did others. In addition, neither gender nor engagement significantly predicted change when considered in the model alone. Only in the context of the other academic, family risk, and mental health covarying factors did these variables significantly predicted change, suggesting that one of the other covarying factors was acting as a suppressor variable. In fact, IQ did act in this manner for both engagement and gender. Possible explanations for this pattern of results include the substantial number of children (particularly girls) at the reported Grade ceiling at age 11, greater decreases in Grades among higher-IQ boys, and a decreasing relationship between academic engagement at age 11 and later Grades caused perhaps by greater stratification of classes by level of difficulty. It is also possible, of course, that systematic changes may occur in engagement over time that we have not measured here. For example, children more highly engaged at age 11 may perhaps be more likely to suffer decreases in engagement as they move through adolescence.

Analyses of Genetic and Environmental Influences

Table 5 shows the MZ and DZ twin correlations for each measure used in our quantitative genetic model fitting, separately for girls and boys. Except for family risk for girls, all of the MZ correlations were higher than the DZ correlations were, and the indicated proportions of genetic influence are shown in the table as well. The magnitudes of the correlations for most of the variables were generally consistent with prior research. All of the correla-

[5] No inherent contradiction exists between this result and the stable or even slightly increasing overall mean Grades shown in Table 2 because the HLM model we used specifically tracked the individual trajectories over time. The means shown in Table 2 reflected the Grades of only those participants contributing data at the specific time points, a series of shifting groups. As noted previously, it was reasonable to consider the unreported data to be missing at random (Little & Rubin, 1987).

Table 4
Significant Standardized Effects and Their p Levels From the Growth Curve Model Indicating Effects on Grades

Parameter	Effect on grades: When estimated alone	*SE*	*p* value	When estimated with other main effects	*SE*	*p* value
Initial grades						
Grand mean	—			2.971	.018	<.001
Gender	.301	.034	<.001	.099	.027	<.001
Engagement	.405	.015	<.001	.283	.015	<.001
IQ	.369	.015	<.001	.286	.012	<.001
Family risk	−.542	.017	<.001	−.067	.012	<.001
Externalizing	−.261	.017	<.001	−.101	.013	<.001
Depression	−.054	.017	.003	—	—	*ns*
IQ × Family Risk	—	—	—	.026	.012	.034
IQ × Externalizing	—	—	—	.025	.012	.032
Linear change in grades						
Grand mean	—			−.172	.045	<.001
Engagement	—	—	*ns*	−.083	.035	.017
Gender	—	—	*ns*	.192	.068	.004

Note. Effects estimated "alone" were those estimates for each covarying factor in the model with only gender, initial grade point average (GPA), covariate, change, change × covariate, and gender × change terms. Effects estimated with others were those estimated for each covarying factor from the full model including all the significant terms. These values reflect the effect on reported Grades of a 1 standard deviation increase in the indicated variable. For change, we show only the covariates significantly predicting it. Gender was coded girls = 1, boys = 0. *ns* = not significant.

tions for the annual reports of average grades at ages 15 and 16 were lower than those for the other ages, probably because of the relative lack of data at these ages. No indication of possible differences was noted in extent of genetic influences between girls and boys. The correlations for family risk deserve some further comment: they were extremely high and differed little between MZs and DZs because many of the specific family risk factors included in the cumulative family risk variable reflected characteristics of the family and parents (race, size of family, parental education, and mental health). Many of these family risk characteristics, however, are under substantial genetic influence in the parental generation (i.e., mental health), which means that genetic and environmental influences will tend to be correlated for the twins for these characteristics. In situations such as this, standard quantitative estimates of genetic, shared and nonshared environmental influences will tend to overstate shared environmental influences and understate genetic influences (Purcell, 2002).

Estimates from the baseline model. Baseline estimates of genetic and environmental influences on initial Grades and linear change in Grades from the initial growth curve model from Figure 1 suggest that the basic pattern of and underlying influences on Grades over time is the same for girls and boys. Confirming this suggestion, we were able to constrain the parameters equal across gender without significant reduction in model fit (change in $-2 \times$ log-likelihood 13.6 on 9 *df*, $p = .14$). This finding did not, however, preclude the possibility that different academic, family risk, or mental health covarying factors would have different effects for the two genders when estimated using the model in Figure 2. In the baseline model, we were also able to constrain genetic and shared environmental influences on residual variances (the latent variables labeled A_{11}–E_{17} in Figure 1) equal over time and across gender (change in $-2 \times$ log-likelihood 36.72 on 26 *df*, $p = .08$). This result indicated little evidence exists for systematic changes over time in the genetic or shared environmental influences on variance in Grades not captured by the linear model.

Both initial Grades and linear change in Grades were under strong genetic influence (the proportion of variance in initial Grades, e.g., genetic can be seen by squaring path a_L in Figure 1; estimate = 70% of variance, 95% confidence interval 56–87%; for change in Grades, it is the square of path a_S in Figure 1, estimate = 70%, 95% confidence interval 55–82%). These genetic influences were only slightly though significantly correlated (path rA Figure 1, .16, 95% confidence interval .01–.31). Shared environmental influences on initial Grades ([path c_L]^2 in Figure 1) totaled 25% (95% confidence interval 8–39%), but were absent from linear change in Grades ([path c_S]^2 in Figure 1). Nonshared environmental influences made up the remaining 5% of variance in initial Grades ([path e_L]^2 in Figure 1, 95% confidence interval 3–6%) and the remaining 30% of variance in change in Grades ([path e_S]^2 in Figure 1, 95% confidence level 18–45%). Ten percent to 15% of the approximately 40% of the total phenotypic variance in Grades that was residual was under genetic influence (squares of paths a_{11}–a_{16} in Figure 1) and 20% to 25% was under shared environmental influence (squares of paths c_{11}–c_{16} in Figure 1). Thus approximately 47% of the total variance in Grades (whether captured by the linear model or not) was under genetic influence, 24% was under shared environmental influence, and the remainder (29%) was under nonshared environmental influence.

Estimates from the extended model. The contributions of genetic and shared and nonshared environmental influences to

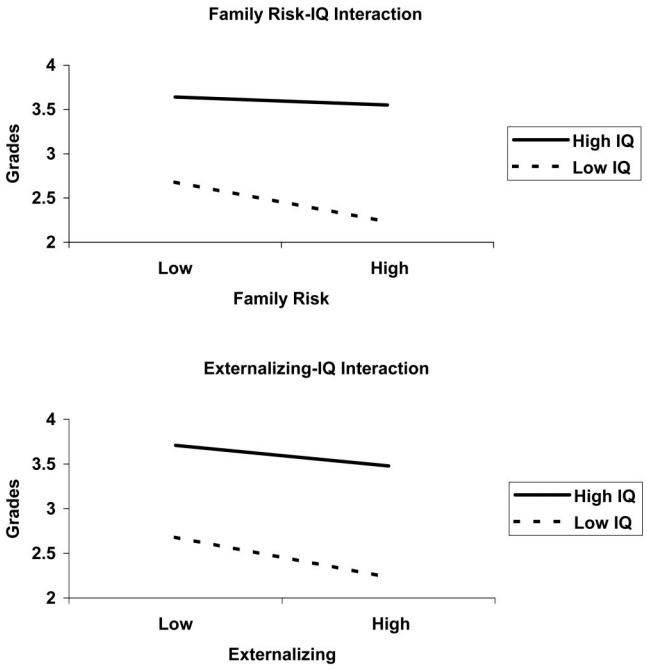

Figure 4. Interactions involving IQ, family risk, and externalizing. Low and High = 2 standard deviations below and above the mean, respectively, based on the HLM regression formula.

Grades, change in Grades, and the academic, family risk, and mental health covarying factors, based on the extended growth curve model in Figure 2, are shown in Table 6. For the Grades variables, the table shows the proportions of total variance, including the residual variance on Grades not captured by the latent growth model as part of total variance in both initial Grades and change in Grades. The basic model with genders estimated separately fit well (χ^2 to saturated model of variances and covariances 66.7, 172 *df*, *p* effectively 1). However, constraining the genetic, shared, and nonshared environmental variances equal across gender (change in $-2 \times$ log-likelihood 529.3 on 63 *df*, *p* effectively 0) was no longer possible; thus Table 6 provides these parameter estimates separately for girls and boys. In general, the estimates of extent of genetic and environmental influence on the individual academic, family risk, and mental health covarying factors were comparable to those from previous research, which acted as a check on the plausibility of our model estimates for which no previous estimates were available. Little evidence existed for differing degrees of genetic and environmental influences for girls and boys, suggesting that the significant deterioration in model fit when the parameters were constrained equal across gender resulted from differing relationships among the predictor variables. Estimates of relationships among genetic and environmental influences of this type are generally somewhat unstable, with large standard errors (Carey & DiLalla, 1994). This instability is complicated by the potential for multicolinearity among the covariates, which can act to increase the instability of our estimates. Nonetheless, description of these relationships in this sample provides valuable information for exploratory purposes.

One way to look at these relationships is to consider the contributions that genetic and environmental influences on each covariate factor make to the genetic and environmental influences on initial Grades and change in Grades. To help this effort, we fit reduced versions of the model shown in Figure 2 involving only one covariate factor at a time. These versions allowed us to estimate the genetic and environmental variance shared with initial Grades and change in Grades by each covariate, disregarding the possibility of multicolinearity with other covariates. Thus, for example, we estimated that 43% of the 51% of variance under genetic influence on initial Grades in girls (see Table 6) might be attributed to genetic influence on engagement, and we estimated that 29% of the 26% of variance under shared environmental influence on initial Grades in boys might be attributed to shared environmental influence on family risk. The full results are shown in Table 7. In addition, we estimated the proportions of genetic and environmental influences on initial Grades and change in Grades that were accounted for in total by the academic, family risk, and mental health covarying factors. Because of the multicolinearity among the genetic and environmental influences on these covarying factors, the sums of their individually estimated contributions exceeded the estimates of the totals; the extent to which this

Table 5
Estimates of Twin Intraclass Correlations and Indicated Genetic Influence for Study Variables

Measure	Girls		Indicated genetic influence	Boys		Indicated genetic influence
	MZ	DZ		MZ	DZ	
Average grades						
Age 11	.81	.55	.52	.84	.56	.56
Age 12	.84	.40	.88	.85	.52	.66
Age 13	.77	.46	.62	.80	.41	.78
Age 14	.80	.55	.50	.80	.59	.42
Age 11 covarying factors						
Age 15	.68	.25	.86	.70	.27	.86
Age 16	.54	.25	.58	.53	.42	.22
Age 17	.77	.41	.72	.74	.47	.54
Engagement	.58	.40	.36	.59	.28	.62
IQ	.78	.53	.50	.73	.56	.34
Family risk	.76	.84	.00	.98	.96	.04
Externalizing	.68	.51	.34	.71	.46	.50

Table 6
Proportions of Phenotypic Variance Attributable to Genetic and Environmental Influences From Extended Latent Growth Model of Grades and Age 11 Covarying Factors

	Girls			Boys		
	A	*C*	*E*	*A*	*C*	*E*
Initial grades	.51 (.30, .69)	.26 (.19, .35)	.23 (.04, .37)	.54 (.30, .75)	.26 (.19, .35)	.20 (.01, .35)
Change in grades	.17 (.00, .57)	.28 (.16, .38)	.55 (.42, .68)	.17 (.00, .58)	.24 (.14, .36)	.58 (.44, .71)
Engagement	.54 (.35, .77)	.09 (.00, .20)	.37 (.33, .42)	.49 (.30, .72)	.10 (.01, .21)	.41 (.37, .46)
IQ	.52 (.34, .76)	.26 (.22, .31)	.22 (.18, .27)	.37 (.18, .60)	.37 (.27, .47)	.26 (.22, .31)
Family risk	.03 (.01, .12)	.77 (.69, .82)	.20 (.16, .25)	.04 (.00, .13)	.94 (.87, .98)	.02 (.00, .19)
Externalizing	.45 (.26, .68)	.25 (.21, .30)	.30 (.26, .35)	.47 (.28, .70)	.25 (.17, .36)	.28 (.24, .33)

Note. *A* refers to proportion of variance attributable to genetic influences, *C* to environmental, and *E* to nonshared environmental influences. 95% confidence intervals are in parentheses.

was true reflected the degree of multicolinearity. As the table indicates, the covarying factors were able to account for most of both the genetic and the shared environmental variance in initial Grades. At the genetic level, the contributions of the covarying factors were relatively independent, particularly for boys. At the shared environmental level, however, substantial overlap was noted; the covarying factors appeared closely intertwined.

Another way to look at these relationships is to consider the proportions of the observed correlations that can be attributed to genetic and shared and nonshared environmental influences. This information is provided in Table 8. For both girls and boys, approximately one half of the moderate correlations between initial Grades and engagement, IQ, and externalizing might be attributed to genetic influences, with the remainder divided roughly between shared and nonshared environmental influences. The correlation between IQ and initial Grades, slightly weaker in boys than it was in girls, also appeared to be under less genetic and greater shared environmental influence in boys than it was in girls. In contrast, the correlation between family risk and initial Grades, again slightly weaker in boys than it was in girls, seemed to be under greater genetic and shared environmental influence in boys than it was in girls. For change in Grades, the great majority of the small observed correlations with the covarying factors was the result of nonshared environmental influences.

Discussion

In this study we investigated several hypotheses and questions involving the factors present at age 11 that contribute to initial Grades and to change in Grades from age 11 to age 17, using a large population-based twin sample. We first investigated the phenotypic relationships, comparing results for girls and boys. Our results were generally consistent with those of other studies that have examined similar questions, but we extend the literature in the phenotypic area with the breadth of factors considered in a single study. We also investigated the extent of, factors contributing to, and relationships among genetic and environmental influences on initial Grades and change in Grades over time, again comparing results for girls and boys. We observed strong genetic influences on all of the variables we considered, though shared

Table 7
Proportions of Genetic and Shared and Nonshared Environmental Influences on Grades Attributable to Individual Age 11 Covarying Factors

	Initial grades			Change in grades		
	A	*C*	*E*	*A*	*C*	*E*
Univariate results						
Engagement	.43	.74	.35	.08	.03	.08
IQ	.34	.56	.15	.05	.00	.04
Family risk	.05	.29	.00	.00	.00	.00
Externalizing: female	.28	.37	.01	.00	.00	.00
Externalizing: male	.04	.92	.04	.00	.00	.00
Multivariate						
Total: female	.86	1.00	.33	.07	.00	.05
Total: male	.74	.97	.54	.09	.03	.04

Note. These proportions reflect the extent to which the age 11 covarying factors accounted for the variance attributable to genetic and environmental influences. Thus, for example, in the multivariate model, 86% of the 51% (Table 6) of genetic variance in initial Grades was captured by the covarying factors. The multivariate total proportion is less than the sum of the univariate parts because of multicolinearity among the univariate parts. Where proportions differed significantly for girls and boys, the proportions are shown separately. As described in the test, these proportions were derived by combining the results of univariate and multivariate analyses. Standard errors of these estimates are not available.

Table 8
Observed Correlations and the Proportions Attributable to Genetic and Shared and Nonshared Environmental Influences

	Change in grades	Engagement	IQ	Family risk	Externalizing
Observed correlations					
Initial grades: girls	.05 (−.06, .16)	.52 (.46, .59)	.54 (.48, .61)	−.29 (−.19, −.39)	−.39 (−.46, −.31)
Change in grades: girls	—	−.03 (−.15, .09)	−.01 (−.13, .11)	−.02 (−.14, .10)	.01 (−.11, .13)
Initial grades: boys	.16 (.06, .26)	.53 (.47, .60)	.45 (.38, .53)	−.19 (−.29, −.09)	−.31 (−.38, −.23)
Change in grades: boys	—	−.01 (−.13, .11)	−.01 (−.13, .11)	−.01 (−.13, .11)	−.02 (−.14, .10)
Genetic					
Initial grades: girls	.56 (.28−.91)	.60 (.32−.95)	.52 (.24−.87)	.26 (−.02, .61)	.68 (.40, 1.00)
Change in grades: girls	—	.10 (−.18, .45)	.00 (−.28, .35)	.08 (−.20, .43)	.09 (−.19, 44)
Initial grades: boys	.98 (.70, 1.00)	.64 (.36, .99)	.37 (.09, .72)	.35 (.07, .70)	.74 (.46, 1.00)
Change in grades: boys	—	.07 (−.21, .42)	.05 (−.23, .40)	.08 (−.20, .43)	.06 (−.22, .41)
Shared environment					
Initial grades: girls	.00 (−.34, .27)	.09 (−.25, .36)	.26 (−.08, .53)	.27 (−.07, .54)	.14 (−.20, .41)
Change in grades: girls	—	.01 (−.33, .28)	.00 (−.34, .27)	.00 (−.34, .27)	.01 (−.33, .28)
Initial grades: boys	.01 (−.33, .28)	.08 (−.26, .35)	.41 (.07, .68)	.43 (.09−.70)	.10 (−.24, .37)
Change in grades: boys	—	.06 (−.28, .33)	.12 (−.22, .39)	.11 (−.23, .38)	.13 (−.21, .40)
Nonshared environment					
Initial grades: girls	.44 (.40, .49)	.31 (.27, .36)	.22 (.18, .27)	.47 (.43, .52)	.18 (.14, .23)
Change in grades: girls	—	.89 (.85, .94)	1.00 (.96, 1.00)	.92 (.88, .97)	.90 (.86, .95)
Initial grades: boys	.01 (−.03, .06)	.28 (.24, .33)	.22 (.18, .27)	.22 (.18, .27)	.16 (.12, .21)
Change in grades: boys	—	.87 (.83, .91)	.83 (.79, .88)	.81 (.77, .86)	.81 (.77, .86)

Note. 95% confidence intervals are in parentheses.

environmental influences were also important. We explore the implications of our observations about genetic and environmental influences following discussion of the phenotypic results.

Phenotypic Results

As we had hypothesized, our HLM model showed significant main effects on initial Grades from most of the age 11 covarying factors we considered: engagement, IQ, family risk, and externalizing behavior (see Table 4). Depression, however, did not have a main effect on initial Grades, independent of the other covarying factors. We observed two-way interactive effects on initial Grades from IQ and family risk and from IQ and externalizing, indicating that high IQ helped reduce the negative effects of family risk and externalizing on Grades. Overall, as we had hypothesized, girls had higher Grades than did boys, and Grades decreased linearly over time somewhat for boys. For girls, an offsetting effect was present such that their Grades increased very slightly over time. In addition, engagement significantly predicted overall change, with higher engagement associated with smaller increases in Grades. The existence of these predictors is important because it shows some systematic patterns in Grades over time that can be captured by a linear equation involving individual- and family-level variables. These systematic patterns can reveal important individual differences in the developmental process associated with moving between the individual's family circumstances and the more commonly shared school environment.

However, large individual fluctuations around the overall rate of change were found, in keeping with the results of Jimerson et al. (1999), and only one of our covarying factors had significant predictive power for change. This finding was primarily because relatively little systematic change was there to predict. By way of comparison, an HLM with no covariates indicated an overall level of initial Grades of 3.02, with individual parameters deviating an average of −0.4 (a small fraction of the mean value) from this. In contrast, the overall average rate of change in Grades was −0.009, with individual parameters deviating an average of −0.064 (a value several times greater than the mean) from this. Both the large individual fluctuations and the failure of most of our academic, family risk, and mental health covarying factors to explain overall change emphasize that the systematic patterns that do exist do not tell the whole story of academic achievement.

In moving from a univariate to a multivariate context, the relative importance of all of our academic, family risk, and mental health covarying factors in predicting initial Grades decreased substantially, as would be expected. Some of the variables maintained their relative importance better, however, than did others. IQ and engagement maintained their levels of relative importance best, indicating that their effects tend to be relatively independent of the effects of the other variables. Engagement and externalizing appeared to explain the majority of gender difference in initial grades, given that the gender effect was reduced from .30 to .10 when the other covariates were added to the model, and these were the two covariates that showed substantial gender differences. In addition, engagement and IQ, and to some extent externalizing, appeared to account almost completely for the effects of family risk because its main effect was reduced from −.54 to −.07 with the addition of the other covariates. This finding has important theoretical implications because it suggests that, at least by age 11, it is not so much the existence of the family risk per se as it is the individual characteristics that are associated with it that have the deleterious effects on school performance. Addressing this possibility will be important for future research designs.

In general, our results regarding main effects corroborated those of Gutman et al. (2003), who authored the other long-term longitudinal study of which we are aware that used HLM to assess achievement trajectories. Our measures differed slightly, given

that we considered engagement in addition to IQ, family risk, and mental health-related behaviors, and we separated externalizing behavior and depression in the evaluation of mental health. In spite of these measurement differences, both studies found that IQ was associated with increased achievement, and cumulative family risk and behaviors indicative of mental health problems were associated with decreased achievement. Unlike Gutman et al. (2003) and others (e.g., Children's Defense Fund, 1993; Entwisle & Alexander, 1992), however, we did not find that greater family risk had an adverse effect on change in achievement over time. This result might be because our sample was more generally representative of the population as a whole (i.e., it was not a high-risk sample). Another possible reason might be that we explicitly measured school engagement, which might capture variance overlapping with family risk. Finally, both studies found interactive effects on achievement between family risk and IQ, and involving mental health.

The nature of the interactive effects, however, differed somewhat in the two studies. In the study by Gutman et al. (2003), greater mental health and higher IQ offered protective effects only to those with relatively low levels of family risk. Thus their interaction provided no explanation for the existence of resilient children (Garmezy, 1993). In contrast, in the current study, higher IQ offered protective effects to children with higher levels of family risk, showing clearly one plausible mechanism through which resilience might develop and be maintained. In the study by Gutman et al. (2003), the mental health–related interaction involved risk; in this study, it involved IQ, but again, the interaction provided a plausible mechanism through which resilience might develop and be maintained. The interactive effects measured in both studies applied only to initial Grades, of course. We have no way of addressing how either set of interactions would hold up throughout the courses of the trajectories.

Given the relative difficulty of replicating interactive effects (Aiken & West, 1991), one important point here may be that the two studies were able to detect interactive effects using similar combinations of variables in samples that probably involved rather different degrees of relative family risk. Though measurement of family risk was not identical, and thus comparison is not straightforward, some comparison is possible. For example, in the current sample, only approximately 9% of the children lived in single-parent families (versus 24% in the study by Gutman and associates), approximately 2% had one or more parents with less than a high-school education (versus 33% of mothers in the study by Gutman's team), and approximately 3% had parents who were unemployed or had laborer or semiskilled jobs (versus 27%). Thus our population-based sample would appear to have included significantly fewer children of really high family risk, and the difference in observed ranges may be the reason for the difference in the nature of the interactive effects.

Another important point in comparing the interactions in the two studies may involve the overall level of achievement in the two studies. Gutman et al. (2003) did not give the average GPA for their sample, but the information they did provide suggests that it was somewhere around 2.1 (their Table 4, page 784). In contrast, average Grades in our sample were in excess of 3.1 (our Table 2). Many possible reasons exist for this difference, but one is that the difference in relative risk in the two samples had with it an associated difference in engagement in school. Engagement in school was not measured in the Gutman study, thus evaluating this possibility was impossible, but it bears investigation in future studies of this nature. Seemingly, the possibility is that, in social, peer, family groups, or any combination in which school is not valued, higher IQ offers no real compensation for relative family risk because higher-IQ students do not invest their greater abilities in school. In social, peer, and family groups in which school is valued, however, higher IQ may do much to compensate for family risk because higher-IQ students do invest their greater abilities in school. Our results involving main effects and our results involving interactions thus both point to the importance of measuring engagement in school along with other variables, indicating individual differences and environmental effects in studies of academic achievement. That engagement in school is important would be consistent with the experience of many immigrant groups to the United States who were uneducated and highly impoverished when they arrived in this country but, within a couple generations, had reached high levels of educational attainment and greatly improved socioeconomic status.

The covarying factors in our model explained initial level of Grades quite well, accounting for 72% of the variance; they did far less well in explaining change in Grades, with only engagement and gender providing any explanatory power. This result is no doubt partly because change in Grades is not systematic and because we measured the covarying factors only at age 11, but it may also be because change is related to variables we did not consider. Obvious candidates are peer relationships and relationships with teachers. Peer relationships might affect school grades in two ways. First, the peer group that an individual adolescent joins may have socializing effects on the adolescent's school performance. Some evidence exists for this notion, even after considering the adolescent's role in selecting the peer group in which he or she participates (Ryan, 2001). Second, when adolescents feel harassed by their peers, the resulting feelings of victimization and low self-worth may affect their school performance. Again, some evidence for this notion can be found (Juvonen, Nishina, & Graham, 2000). As peer groups in adolescence tend to be rather unstable (Ryan, 2001), shifting peer relationships may help explain the otherwise unsystematic changes in Grades. Relationships with teachers may also affect reported Grades because of differing grading standards, differing degrees of teaching ability, and differing responses to their teaching methods from their students. The individual characteristic most likely to be mediated by these social factors would appear to be engagement in school, and, measured at age 11, this factor was the one in our model that did predict change, though in a counterintuitive way. The ways in which social factors affect achievement trajectories deserves further attention in future research.

Results Involving Genetic and Environmental Influences

In this study, we also investigated the extent of genetic and environmental influences on initial Grades and linear change in Grades, as well as their genetic and environmental relations with the academic, family risk, and mental health covarying factors. In considering our findings in this area, important points to keep in mind are that (1) our model measured only the genetic and environmental relationships among the academic, family risk, and mental health covarying factors and *predictable* linear change and

that (2) systematic linear change of this type was a relatively small part of the overall change that took place over time. That is, as in other studies (e.g., Jimerson et al., 1999; Kowalski-Jones & Duncan, 1999), we observed substantial individual-level change that was not systematic in this sense, though some of the individual-level change that could not be so captured appeared to be under genetic and shared environmental influence.

We observed strong genetic influences and moderate shared environmental influences on level of achievement. Given the substantial main effects of the academic, family risk, and mental health covarying factors on initial Grades and the well-established estimates of genetic and shared environmental influences on these covarying factors, this observation was not surprising. Interestingly, however, we also observed genetic influences on linear change in achievement, though these genetic influences were not linked with the covarying factors as measured at age 11. The presence of these genetic influences on linear change implies some biological involvement exists in whatever developmental mechanisms operate during adolescence to lead to these patterns of academic achievement. Clearly, many genetically influenced biological mechanisms are associated with physical puberty that operate during adolescence, some of which might be associated with changes in achievement, but it is also possible that genetically influenced developmental psychological mechanisms exist that operate similarly. Obvious candidates for such a role would include whatever genes are involved in the increased expression of externalizing behaviors and depression during adolescence (Eaves & Silberg, 2003; Moffitt, 1993; Silberg et al., 1999), but another would be genes involved in cognitive maturation (Giedd et al., 1999). Additional evidence was found for shared environmental influence on overall change in achievement, possibly reflecting changing parental expectations for academic achievement during adolescence, as well as grading policies that differed from school to school.

In the basic model of genetic and environmental influences, we were able to constrain the parameters equal for girls and boys, implying that the mean gender differences in achievement that have been observed do not result from gender differences in the overall pattern of genetic and environmental influences on academic achievement over time. This conclusion is consistent with the observation that, in the phenotypic model, the overall gender difference in Grades (coefficient of .3 from Table 4) was largely explained by the other covarying factors (coefficient of .099 in the full model). In addition, gender differences were not readily apparent in the proportions of genetic and environmental influences on the academic, family risk, and mental health covarying factors. This did not mean, however, that the relations among these covarying factors and achievement would not show gender differences.

Our extended model did show some differences between girls and boys, though the overall patterns of results were similar. For both genders, the observed correlations among engagement, IQ, and externalizing were largely genetically mediated. The proportions of genetic influence contributed by these covarying factors generally reflected these observations. In total, our covarying factors accounted for approximately 80% of the genetic influence on initial Grades for both genders. For both girls and boys, the relationships among engagement, IQ, and family risk were stronger at the level of the shared environment than they were at the genetic level. This finding is indicated by the higher proportions of shared environmental variance than genetic variance explained by the covariates in Table 7 and suggests that shared environmental influences act to reinforce and draw together genetic influences. This result was true to a greater degree in boys than it was in girls; it also requires further investigation, especially because assortative mating for academic achievement may overstate estimates of shared environmental influences and understate estimates of genetic influences (Falconer & Mackay, 1996).

This study is subject to some methodological limitations that should be considered when evaluating the significance and generalizability of our results. First, our assessment of behaviors and achievement is based on children's, parents', and teachers' reports rather than direct observation or actual report cards of grades from a consistently administered system. With respect to the grades reports, we did not take into consideration the possibility that the grades were earned in accelerated or remedial classes. Our composite reporter measure, however, correlated very highly with actual grade transcripts in a subsample for which such transcripts were available, and the lack of recognition of accelerated and remedial classes should have served to blunt our ability to detect associations of grades with our academic, family risk, and mental health covarying factors. Second, our sample is predominantly Caucasian and includes relatively few people in abject poverty, given that it is representative of twin births in Minnesota for the birth years in question. This type of sampling is an advantage in the sense that it may help clarify relationships within this broad-based population group, but the generalizability of our findings to other ethnic groups and more extreme samples needs to be addressed. Nevertheless, the study presents a clear picture of the factors affecting trajectories of academic achievement and their relative importance in this sample. In addition, the results make clear that genetic influences on these factors are important in explaining individual differences in these trajectories. At the same time, shared environmental influences are also important and may be more tightly linked, thus serving to reinforce each other more directly. We need to understand this relationship more thoroughly, given that it may have significant implications for the development of prevention and intervention programs intended to maximize each child's achievement level.

As we noted in describing our analytic strategy for this study, the methods we used to estimate proportions of genetic and environmental influences are based on the assumption that genetic and environmental influences are independent. The independence assumption implies that no gene-environment interactions or correlations that would act to create differing degrees of genetic and environmental influences exist within different subgroups of the sample. Violations of this assumption do not, however, invalidate the overall approach. Rather, they render the estimates applicable only on an overall average population–level basis, and they introduce systematic distortions in the estimates. These distortions have different effects, depending on the nature of the interaction or correlation. Specifically, interaction between genetic and shared environmental influences acts to increase the proportion of genetic influence; interaction between genetic and nonshared environmental influences acts to increase the proportion of nonshared environmental influence. Correlation between genetic and shared environmental influences acts to increase the proportion of shared environmental influence; correlation between genetic and nonshared environmental influences acts to increase the proportion of

genetic influence (Purcell, 2002). By their nature, gene-environment interactions and correlations are multivariate. Thus the first step in addressing the possibility of their existence is exactly the one we followed in this study: rather than estimating proportions of genetic and environmental influences on individual variables taken one at a time, we estimated these proportions in a multivariate context in which we might get some sense of the interrelations that may be involved and in which any peculiarities in the covariances for one variable might be mathematically smoothed by the presence in the model of the other variables. The associations in this study showing strong genetic or shared environmental mediation (or both) are prime candidates for exploring gene-environment interaction and correlation in subsequent research.

Several recent studies have raised another issue affecting the interpretation of data from twin studies: the variance attributed to genetic sources and usually interpreted as arising from genetic differences among people in studies such as this one may arise instead from environmental influences that lead to differential expression of genes that do not differ among people. For example, Weaver et al. (2004) reported that differential maternal treatment of rat pups in infancy resulted in differential expression of glucocorticoid receptor genes that, in turn, resulted in differential response to stress that persisted into adulthood. Naturally, twin studies cannot distinguish strictly genetic from epigenetic sources of variance, and epigenetic sources of variance are likely to be of great importance in explaining the development of psychological processes and patterns of behavior. However, that this inability necessarily produces confusion in interpreting only the genetic sources of variance is not true. The maternal treatment that led to the differences in genetic expression in the Weaver et al. (2004) study, for example, would, in a twin-study context, contribute to shared environmental rather than genetic variance because presumably all pups reared by the same mother would experience the same relative levels of maternal licking and grooming. In addition, the expression of particular genes is often under the control of other genes, and, in the aggregate, the regulation of genetic expression appears to be substantially under genetic influence (York et al., 2005). Twin studies reflect the degree to which the entire genome influences the outcome in question.

Developmentalists often note a related problem with estimates of proportions of genetic and environmental influence on behavioral characteristics such as intelligence, school engagement, family risk, mental health, and personality: known single nucleotide polymorphisms and even polymorphic genetic regulatory networks account at best for negligible proportions of variance. Genetic epidemiologists face this problem as well when searching for the specific genes involved in medical conditions such as heart disease and diabetes that are generally recognized to be genetically influenced (Deery, 2000; Horenstein & Shuldiner, 2004). In addition, we know as little about the specific genes involved in height (Willemsen et al., 2004) as we do about those involved in intelligence, yet people routinely estimate how tall children will become based on the heights of their parents. Inability to identify the specific genes involved in a trait does not mean that no genetic involvement exists. Rather, it means that we have not yet developed the techniques that will ultimately help us understand this involvement. The techniques we used in this study are only very preliminary steps in that process. Nonetheless, they indicate the importance of both genetic and environmental influences to reported school grades, and they suggest that shared environmental influences act to reinforce and draw together relatively independent genetic influences. This information is important for future research.

In recent articles and commentaries in *Developmental Psychology,* several authors have debated the value of behavioral genetic methods for understanding parent-adolescent relationships (Greenberg, 2005; McGue, Elkins, Walden, & Iacono, 2005a, 2005b; Partridge, 2005). Given the currency of this debate and our use of behavior genetic methods to examine a different aspect of adolescent development, considering our findings in light of the points raised in these papers is useful. First, we make no claim that the genetic influences identified in this study imply that genes determine behavioral outcomes. Rather, studies such as this one reveal that, in the environments experienced by the individuals in these samples, variation in genetic expression (whether from genomic variation or variation in expression of the same genes caused by the action of other genes) accounts for some portion of the variance across individuals in behavioral outcomes. This idea is a broad statement of the manifestation of genetic influences on individual differences that has nothing to do with the identification of the specific genes and genetic effects that underlie the statistical variation. Such manifestation of genetic influences does, however, suggest that something intrinsic to the individual is of critical importance in the emergence of behavior. In a multivariate context such as that applied in this study, we also learn that variance in genetic expression contributing to one behavior such as school grades earned is more or less related to genetic variance contributing to other traits such as expressed engagement in school, disruptive behavior, and IQ. This foundation helps articulate the individual characteristics that are important in the genetic and environmental transactions that lead to the behavior.

Second, statistical models are simplifications of actual psychological processes that rely on underlying assumptions about specific aspects of these processes, and all statistical models are applied in situations in which some or all of the underlying assumptions may be violated. Quantitative behavior genetic models are no exception to this rule, as we have tried to stress in discussing the limitations to this study. This understanding does not mean, however, that these models have nothing to tell us about the etiology of human behavior. Rather, it means that their results should be evaluated in light of their limitations, and that other models that rely on other assumptions should be applied and their results and limitations evaluated as well. We are currently unable to specify the dynamic genetic and environmental transactions and the associated neurological pathways presumed to underlie behavior. Nevertheless, research such as that reported here brings us a little closer to achieving this valuable goal because it helps articulate patterns of genetic activity in the environments in which our study participants live.

References

Ackerman, B., Izard, C., Schoff, K., Youngstrom, E. A., & Kogos, J. (1999). Contextual risk, caregiver emotionality, and the problem behaviors of six- and seven-year-old children from economically disadvantaged families. *Child Development, 70,* 1415–1427.

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage Publications.

American Association of University Women. (1996). *Gender gaps.* New York: Marlowe.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.

Bouchard, T. J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212,* 1055–1059.

Bronfenbrenner, U., McClelland, P., Wethington, E., Moen, P., & Ceci, S. J. (1996). *The state of Americans.* New York: Free Press.

Burchinal, M. R., Roberts, J. E., Hooper, S., & Zeisel, S. A. (2000). Cumulative risk and early cognitive development: A comparison of statistical risk models. *Developmental Psychology, 36,* 793–807.

Butler, S. R., Marsh, H. W., Sheppard, M. J., & Sheppard, J. L. (1985). Seven-year longitudinal study of the early prediction of reading achievement. *Journal of Educational Psychology, 77,* 349–361.

Butts, J. A., Snyder, H. N., Finnegan, T. A., Aughenbaugh, A. L., Tierney, N. J., Sullivan, D. P., et al. (1995). *Juvenile Court Statistics: 1992.* Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Carey, G., & DiLalla, D. L. (1994). Personality and psychopathology: Genetic perspectives. *Journal of Abnormal Psychology, 103,* 32–43.

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist, 52*(10), 1051–1058.

Children's Defense Fund. (1993). *Progress and peril: Black children in America.* Washington, DC: Author.

Cohen, P., Cohen, J., Kasen, S., Velez, C. N., Hartmark, C., Johnson, J., et al. (1993). An epidemiological study of disorders in late childhood and adolescence: 1. Age and gender-specific prevalence. *Journal of Child Psychology and Psychiatry, 34,* 851–867.

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish Mental Survey of 1932. *Intelligence, 31,* 533–542.

Deater-Deckard, K., Dodge, K. A., Bates, E. A., & Pettit, G. S. (1998). Multiple risk factors in the development of externalizing behavior problems: Group and individual differences. *Development and Psychopathology, 10,* 469–493.

Deery, T. A. (2000). The evolving role of genetics in the diagnosis and management of heart disease. *Nursing Clinics of North America, 35,* 963.

DuBois, D. L., Eitel, S. K., & Felner, R. D. (1994). Effects of family environment and parent-child relationships on school adjustment during the transition to early adolescence. *Journal of Marriage and the Family, 56,* 405–414.

Eaves, L., & Silberg, J. (2003). Modulation of gene expression by genetic and environmental heterogeneity in timing of a developmental milestone. *Behavior Genetics, 33,* 1–6.

Eccles, J. S. (1984). Sex differences in achievement patterns. In T. Sonderegger (Ed.), *Nebraska Symposium on Motivation* (pp. 97–132). Lincoln, NE: University of Nebraska Press.

Eccles, J. S., Roeser, R., Wigfield, A., & Freedman-Doan, C. (1999). Academic and motivational pathways through middle childhood. In L. Balter & C. S. Tamis-LeMonda (Eds.), *Child psychology: A handbook of contemporary issues* (pp. 287–317). Philadelphia: Psychology Press.

Entwisle, D., & Alexander, K. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review, 57,* 72–84.

Falconer, D. S., & Mackay, T. F. C. (1996). *Quantitative genetics* (4th ed.). Harlow, England: Prentice Hall.

Fergusson, D. M., & Horwood, L. J. (1995). Early disruptive behavior, IQ and later school achievement and delinquent behavior. *Journal of Abnormal Child Psychology, 23,* 183–199.

Fergusson, D. M., Horwood, L. J., & Lynsky, M. T. (1993). The effects of conduct disorder and attention deficit in middle childhood on offending and scholastic ability at age 13. *Journal of Child Psychology and Psychiatry, 34,* 899–916.

Frick, P. J., Kamphaus, R. W., Lahey, B. B., Loeber, R., Christ, M. A. G., Hart, E. L., & Tannenbaum, L. E. (1991). Academic underachievement and the disruptive behavior disorders. *Journal of Consulting and Clinical Psychology, 59,* 289–294.

Garmezy, N. (1993). Children in poverty: Resilience despite risk. *Psychiatry, 56,* 127–136.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience, 2*(10), 861–863.

Gomez, R., Harvey, J., Quick, C., Sharer, I., & Harris, G. (1999). *DSM–IV* AD/HD: Confirmatory factor models, prevalence, and gender and age differences based on parent and teacher ratings of Australian primary school children. *Journal of Child Psychology and Psychiatry, 40,* 265–274.

Greenberg, G. (2005). The limitations of behavior genetic analyses. *Developmental Psychology, 41,* 827–832.

Gutman, L. M., Sameroff, A. J., & Cole, R. (2003). Academic growth curve trajectories from 1st grade to 12th grade: Effects of multiple social risk factors and preschool child factors. *Developmental Psychology, 39,* 777–790.

Gutman, L. M., Sameroff, A. J., & Eccles, J. S. (2002). The academic achievement of African American students during early adolescence: An examination of multiple risk, promotive, and protective factors. *American Journal of Community Psychology, 30*(3), 367–399.

Hedges, L. V., & Nowell, A. (1999). Changes in the black-white gap in achievement test scores. *Sociology of Education, 72*(2), 111–135.

Hetherington, E. M. (1989). Coping with family transitions: Winners, losers, and survivors. *Child Development, 60,* 1–14.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111,* 127–155.

Hollingshead, A. B. (1957). *Two factor index of social position.* New Haven, CT: August B. Hollingshead.

Horenstein, R. M., & Shuldiner, A. R. (2004). Genetics of diabetes. *Reviews in Endocrine and Metabolic Disorders, 5,* 25–36.

Hyde, J. S., & Kling, K. C. (2001). Women, motivation, and achievement. *Psychology of Women Quarterly, 25,* 364–378.

Iacono, W. G., Carlson, S. R., Taylor, J., Elkins, I. J., & McGue, M. (1999). Behavioral disinhibition and the development of substance-use disorders: Findings from the Minnesota Twin Family Study. *Development and Psychopathology, 11,* 869–900.

Jöreskog, K., & Sörbom, D. (2002). *LISREL 8.53: User's reference guide.* Chicago: Scientific Software International.

Jimerson, S., Egeland, B., & Teo, A. (1999). A longitudinal study of achievement trajectories: Factors associated with change. *Journal of Educational Psychology, 91,* 116–126.

Jockin, V., McGue, M., & Lykken, D. T. (1996). Personality and divorce: A genetic analysis. *Journal of Personality and Social Psychology, 71*(2), 288–299.

Johnson, W., McGue, M., & Iacono, W. G. (2005). Disruptive behavior and school grades: Genetic and environmental relations in 11-year-olds. *Journal of Educational Psychology, 97,* 391–405.

Juvonen, J., Nishina, A., & Graham, S. (2000). Peer harassment, psychological adjustment, and school functioning in early adolescence. *Journal of Educational Psychology, 92,* 349–359.

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105,* 198–214.

Kowalski-Jones, L., & Duncan, G. J. (1999). The structure of achievement and behavior across middle childhood. *Child Development, 70,* 930–943.

Krueger, R. F. (1999). The structure of common mental disorders. *Archives of General Psychiatry, 56,* 921–926.

Krueger, R. F., Caspi, A., & Moffitt, T. (2000). Epidemiological personology: The unifying role of personality in population-based research on problem behaviors. *Journal of Personality, 68*(6), 967–997.