sociological science

Gender Typicality and Academic Achievement among American High School Students

Jill E. Yavorsky,[a] Claudia Buchmann[b]

a) University of North Carolina Charlotte; b) The Ohio State University

Abstract: This study is the first to use nationally representative data to examine whether differences in gender-typical behaviors among adolescents are associated with high school academic performance and whether such associations vary by race or socioeconomic status. Using wave I data from the National Longitudinal Study of Adolescent Health and linked academic transcript data from the Adolescent Health and Academic Achievement study, we find that boys who report moderate levels of gender atypicality earn the highest grade point averages (GPAs), but few boys score in this range. As gender typicality increases, boys' GPAs decline steeply. In contrast, girls who practice moderate levels of gender typicality earn slightly higher GPAs than other girls. These patterns generally hold across race and socioeconomic status groups.

Keywords: achievement; femininity; gender; gender typicality; high school students; masculinity

A significant body of research has found that students' academic performance in middle and high school strongly predicts their later academic attainment and other life outcomes. Academic performance in high school, typically measured by grade point average (GPA), is positively associated with college enrollment and completion, and students with good grades in high school are much more likely to obtain a college degree than poorly performing students (Bowen, Chingos, and McPherson 2009; Roderick, Coca, and Nagaoka 2011). Buchmann and DiPrete (2006) demonstrate that gender differences in academic performance prior to college largely explain the growing female-favorable gender gap in college completion; they maintain that high school grades are a better predictor than standardized test scores of future academic success because grades are indicative of behaviors learned early in life that tend to persist into high school and college (DiPrete and Buchmann 2013). These large average differences between boys and girls in behaviors critical to later educational success (e.g., completing homework, not skipping school, bringing materials to class) may be due, in part, to societal gender norms. For example, some ethnographic research finds that adolescent masculinity norms are antithetical to academic engagement and effort (Willis 1977; Morris 2008) and therefore, some boys enact masculinity by valorizing physical skills, taking risks, and ignoring authority figures rather than doing well in school.

Of course, there is great variation in the degree to which students adhere to societal gender norms. Here the concept of "doing gender" is useful because it reminds us that although gender is often presented as a binary characteristic (and measured as such in educational research), it is multidimensional—something people do and practice rather than something they are (West and Zimmerman 1987). Students of the same gender vary considerably in their practice of gender-typical behaviors, or those behaviors generally more associated with one sex than the

Citation: Yavorsky, Jill E., and Claudia Buchmann. 2019. "Gender Typicality and Academic Achievement among American High School Students." Sociological Science 6: 661-683.

Received: October 8, 2019

Accepted: November 13, 2019

Published: December 12, 2019

Editor(s): Jesper Sørensen, Kim Weeden

DOI: 10.15195/v6.a25

other. Research has linked the degree to which adolescents practice gender-typical behaviors to a wide range of outcomes, including sexual orientation (Lippa 2000; Udry and Chantala 2004), psychological well-being (Kreiger and Dumka 2006), relationship pairing (Udry and Chantala 2004), and occupational choice (Ueno, Roach, and Peña-Talamantes 2013). Scholars have also found associations between students' gender typicality and their school engagement and attachment (Ueno and McWilliams 2010) as well as their standardized test scores (Lippa 1998). To the best of our knowledge, however, no research to date has utilized nationally representative data to examine whether differences in gender-typical behaviors among boys and among girls are associated with academic performance in high school or whether such associations, if they exist, vary by race or socioeconomic status. Some studies have examined the role masculinity plays in boys' lower academic performance, and although important theoretical advancements stem from this work, these studies tend to be qualitative investigations of small and nonrepresentative (and often non–United States) samples (Jackson and Dempster 2009; Morris 2012; Archer et al. 2012). Moreover, little attention has been devoted to examining girls' gender expressions and how they relate to academic performance. Given the wide variation in girls' academic performance, a focus on their gendered practices is also critical.

This study uses nationally representative data—the National Longitudinal Study of Adolescent Health (Add Health) gathered in the mid-1990s—to examine how gender-typical behaviors relate to high school students' academic performance. It addresses several questions: 1) What is the distribution of gender-typical behaviors within the U.S. population of adolescent boys and girls? Does the distribution of gender typicality among boys and among girls vary by race and socioeconomic status? 2) How is the gender typicality of boys and girls related to their subsequent overall high school GPA? And how is it related to GPA in English and math, subjects that are stereotypically viewed as feminine and masculine, respectively? To address these questions, we adapt a scale from Cleveland, Udry, and Chantala (2001) to measure the degree to which respondents practice a set of behaviors typical of the average boy or average girl in the sample. We use this gender typicality scale as a proxy for normative masculine and feminine behavior on the basis that some behaviors (i.e., risk taking, crying) become socially attached to the sex that more commonly displays them. We acknowledge that societal gender norms significantly influence the frequency of these behaviors, thus this scale measures the degree of conformity to norms and practices that are more or less characteristic of girls and boys; it makes no claims about biological determinacy of these characteristics.

Gendered practices change over time, and our study addresses the above questions for a cohort who came of age in the 1990s. To the best of our knowledge, there is not a more recent, nationally representative data set that includes transcript data to address this issue, but an examination of this cohort of U.S. youth, who are now working-age adults, is valuable for several reasons. First, although gender norms have expanded in recent decades (particularly for girls and given recent queer and transgender movements), evidence indicates that most adolescents continue to abide by stereotypical gender norms (Chu, Porche, and Tolman 2005; Pascoe 2007; Archer et al. 2012). Second, the analysis of the potential relationship between

gender typicality and academic performance of this cohort provides a valuable baseline that, should more current data become available, could be compared with data for a more recent sample of adolescents in order to assess the degree of change in gendered behaviors and their relationship to academic outcomes.

Theoretical Perspectives on Gender Practices and Academic Performance

Although U.S. research on gender and education sometimes acknowledges the heterogeneity within the population of boys and the population of girls (Diprete and Buchmann 2013:14), it generally focuses on how differences in the average behaviors and orientations of boys versus girls are related to gender gaps in academic performance and educational attainment. Using nationally representative data over four decades, DiPrete and Buchmann (2013:86) find a statistically significant, female-favorable gap in overall high school GPA for each time point (1972, 1982, 1992, 2004) of roughly 0.30 on a 4.0 scale. They maintain that much of this gender gap in GPA is due to lower levels of school engagement and performance-related behaviors among boys. We expand the study of gender and education from a binary focus on differences between girls and boys to examine gender practices among boys and among girls as they relate to academic performance.

Generally, people strive to practice gender in ways that increase their status in a given context (West and Zimmerman 1987). For boys, this often means behaving in ways typically associated with the male sex, such as enacting hegemonic masculinity (Connell 2005). Although debate about what is considered hegemonic continues, scholars generally agree that core characteristics of masculinity include dominance, authority, athleticism, constrained emotional expression, mental and physical strength, rationality, and sexual prowess (Smiler and Epstein 2010; Legewie and DiPrete 2012). Idealized forms of masculinity influence how boys act and define their social position among peers (Schrock and Schwalbe 2009). Accordingly, some behaviors, such as fighting and risk taking become more typical of one gender than the other. Qualitative research suggests that boys who behave in ways more typical of their gender may be at risk for lower academic performance (Jackson and Dempster 2009; Morris 2012). Participating in masculine norms is not only about the behaviors that boys do practice, but also about those behaviors that boys do not practice, such as those associated with femininity. Boys who either self-identify as "nerds" or have been deemed "nerds" by peers may engage in behaviors more typically associated with girls and femininity, such as cooperativeness and attentiveness (Heyder and Kessels 2013), and these boys may earn higher GPAs than boys who engage stringently in hegemonic masculine behaviors.

Although norms of masculinity have remained relatively narrow (England 2010), gender norms for girls and women have broadened beyond the stereotypical feminine traits of congeniality and dependence to include characteristics like assertiveness and independence (Seem and Clark 2006). These changes may serve their academic achievement because girls can exhibit behaviors historically associated with males, such as competence and intelligence, in conjunction with traditionally

feminine behaviors like cooperativeness. Although practicing both masculinities and femininities may be related to higher academic achievement among girls, those who are highly gender typical may earn lower GPAs if they feel pressure to focus on their appearance and relationships instead of their academic performance, as some qualitative non-U.S. research intimates (Archer, Halsall, and Hollingworth 2007; Archer et al. 2012).

Intersectionality and Gender Practices

Depending on their position in racial and class hierarchies, adolescents may display different gendered practices in order to maximize their access to resources and signal status (Collins 2000; Connell and Messerschmidt 2005). Schrock and Schwalbe (2009) explain that "manhood acts are aimed at claiming [gender-based] privilege, eliciting deference, and resisting exploitation" (P. 281). Acts that are interpreted as masculine vary "depending on other features of the actor (age, race, ethnicity, class), the audience, and the situation" (P. 280). For example, some economically disadvantaged men use "compensatory manhood acts" such as fighting and violence to reassert their masculinity and dominance in the face of economic circumstances perceived as emasculating (Pyke 1996). Likewise, when school personnel stereotype black and Hispanic boys as unruly and requiring greater discipline or surveillance (Ferguson et al. 2000), they may break the rules, talk back to teachers, and disdain academic achievement in order to resist exploitation and elicit deference from peers (Willis 1977; MacLeod 2008). "The lesson – for boys who are marginalized because of class or race – is that a masculine self can be signified, and deference elicited, by evoking fear in others" (Schrock and Schwalbe 2009:283). On the other hand, Asian American youth may benefit academically from the tendency to be stereotyped as academically talented ("stereotype promise") as well as having less rigid conceptions of masculinity among Asian Americans more generally. Even so, Hsin (2018) finds that these protective factors erode during adolescence, such that Asian American boys underperform relative to their female counterparts in high school.

The implications for socioeconomic status and race are less straightforward for girls' GPAs. On one hand, black girls may have greater leeway than white girls to practice multifaceted gendered behaviors (including masculine-typed behaviors like assertiveness and authority) (Morris 2007). On the other hand, they may rely on risky or illicit behaviors to gain status, due partly to an inability—regardless of how they behave—to gain cultural feminine status because this status is inextricably tied to whiteness (Livingston, Rosette, and Washington 2012). Lower class girls may also practice a tough persona, like that often associated with boys (Brown 1997). Conversely, they may adhere to more traditional feminine norms given that less-educated parents often hold more traditional gender attitudes and encourage such norms with their daughters (Davis and Greenstein 2009). To the extent that emphasized feminine behaviors are associated with lower GPAs, the relationship may be particularly strong among girls of lower socioeconomic status. Other research finds that among Asian Americans, traditional gender norms lead to high parental expectations and the tendency of immigrant parents to exert greater control over their daughters than their sons and these behaviors are beneficial

to girls' academic performance (Qin 2009; Hsin 2018). Thus, girls and boys may practice gender differently depending on their location within the stratification hierarchy, and these differences may be related to their academic achievement.

Gender Practices Measured by Gender Typicality Scales

Scales based on gender-diagnostic (GD) techniques provide a compelling way to measure gender norms that stem from an array of cultural, social, and historical factors (West and Zimmerman 1987). Grounded in methodologies developed by Terman and Miles (1936) and Lippa and Connelly (1990), GD techniques take a Bayesian approach to evaluate a broad range of items in a particular questionnaire and pinpoint those that predict respondents' genders. Specifically, GD scales assess the probability that an individual is male or female based on how they answer a series of questionnaire items compared to the averages for males or females in the sample. Although prior scales have largely been based on perceptions of males' and females' desired traits of each sex (Bem 1974; Auster and Ohm 2000), the GD scale is determined by actual reported behavioral differences that distinguish males' typical responses from females' typical responses for particular items. Although no scale can capture the manifold ways that boys and girls do gender, prior quantitative research has used scales derived from GD techniques to examine how different expressions of gender relate to school outcomes, such as academic motivation (Ueno and McWilliams 2010) and standardized test scores (Lippa 1998).

The current study examines associations between common gender practices and high school academic performance at the population level. We expect to find that gender-atypical boys—those who practice fewer masculine-typed behaviors and more feminine-typed behaviors—earn higher overall GPAs relative to other boys. Gender typicality may be less related to boys' GPA in math than their GPA in English, as strong performance in female-typed subjects may weaken their masculine identity and status (Leaper, Farkas, and Brown 2012). We expect to find less variation in the relationship between gender typicality and GPA across school subjects for girls. We also examine social class and racial variations in these relationships.

Methods

Data and Sample

Analyses are based on wave I in-home data of the National Longitudinal Study of Adolescent Health (Add Health) and wave III academic transcript data of the Adolescent Health and Academic Achievement (AHAA) study—an expanded Add Health survey that includes additional academic achievement measures linked to 20,745 adolescents who were interviewed in 1994 and 1995 (along with 17,700 parents).[1] The study includes oversamples based on race, twin and disability status, and other categories. The AHAA study, conducted between July 2001 and April 2002, provides academic transcripts for 7th through 12th grade for wave III sample respondents who also completed the wave I survey six years earlier. It leveraged the

Classification of Secondary School Curriculum to code and standardize courses into general classes, such as math and English.[2] This data set is used to construct the dependent variables: overall GPA, English GPA, and math GPA. The six-year lag between wave I and wave III provides time for survey respondents to complete high school (unless they drop out) even if they were in seventh grade when wave I was administered. Accordingly, we are able to construct GPA values using transcript data from all years that respondents attended high school.

The sampling frame includes boys and girls enrolled in grades 9 through 12 in 1994 through 1995. We focus on high school students because most youth have gone through puberty by this point and pressure to conform to gender norms is strong during this period (Simmons et al. 1987). Of the 15,170 respondents who participated in wave III of the Add Health survey, the AHHA collected transcript data for approximately 12,000.[3] From this sample, we drop all seventh and eighth grade students (3,100) and students with missing grades on their transcripts (200) to produce a sampling frame of 8,700 9th- through 12th-grade students. We used listwise deletion to address missing observations for control variables (Rubin 1996; Graham, Olchowski, and Gilreath 2007) to obtain unbiased estimates of population parameters and standard errors.[4] This yields an analytical sample of 7,224 respondents.

Measures

Gender typicality. We use data from wave I to capture adolescents' self-reported behaviors during high school. We adapt a GD scale created by Cleveland et al. (2001) based on Add Health wave II data and later used by Udry and Chantala (2004) and Ueno and McWilliams (2010). From an extensive range of 50 items that showed moderate sex differences, Cleveland et al. used a stepwise procedure to narrow this list to 16 items that strongly predict respondents' sex, thus providing an internally valid and reliable measure of behaviors typical of one sex or the other (Udry and Chantala 2004; Ueno, Roach, and Peña-Talamantes 2013). Wave I data include 12 of the 16 questions originally used by Cleveland et al. (2001).[5] By adapting this scale to wave I data, we were able to capture respondents' self-reported behaviors during the time they earned high school grades. See Section A1 of the online supplement for additional details on robustness checks that affirm the similarity of our scale to prior scales.

Using logistic regression, the gender typicality measure considers participants' responses to 12 items to produce a predicted value that a respondent is male (1) or female (0). Following prior research, responses to each question are equally weighted in the scale (i.e., no one item is more heavily weighted than another item in determining a respondent's overall gender typicality score) (Udry and Chantala 2004; Ueno, Roach, and Peña-Talamantes 2013). Table 1 reports logistic regression coefficients for each item in the scale. From the full analytic sample of boys and girls, we constructed a gender typicality scale for boys ranging from 0.01 to 0.99, with the latter indicating that the respondent has a predicted probability of 0.99 of being male and engages in behaviors predominately practiced by other boys and very few girls. The gender typicality scale for girls is derived from the same 12 questions but with

Table 1: Questionnaire items for the gender typicality scale.

Item	Regression Coefficients	Direction of Response Pattern	(Boy M - Girl M) / (Boys' SD)	Boys' SD
Frequency of crying	−1.47	0 = never, 4 = everyday	−0.90	0.56
Frequency of moodiness	−0.15	0 = never, 4 = everyday	−0.15	2.53
Frequency of poor appetite	−0.20	0 = never, 4 = everyday	−0.17	1.77
How honestly answered questions	0.32	1 = completely honestly, 4 = not honestly at all	0.08	1.78
Trouble paying attention	0.40	0 = never, 4 = everyday	0.07	2.73
Bothered by things	−0.09	0 = never, 3 = most or all of the time	−0.13	1.11
How physically fit	0.37	0 = strongly disagree, 4 = strongly agree	0.24	2.00
Past 12 months of frequency of serious fighting	0.67	0 = never, 3 = five or more times	0.13	1.94
Frequency of exercising	−0.11	0 = not at all, 3 = five or more times	0.01	2.37
Frequency of cycling/skate boarding	0.29	0 = not at all, 3 = five or more times	0.10	1.97
Do you like yourself as you are	0.19	0 = strongly disagree, 4 = strongly agree	0.19	2.12
Upset by problems	0.21	1 = strongly agree, 5 = strongly disagree	0.14	2.15

Notes: Total N = 7,224; boys = 3,455; girls = 3,769. Using logistic regression and the predict function in Stata 14, the gender typicality measure considers participants' responses to each of the above 12 questions to produce a predicted value that a respondent is a boy or girl.

scores reversed, such that it ranges from 0.01 to 0.99, with the latter indicating that the respondent has a predicted probability of 0.99 of being female and engages in behaviors predominately practiced by other girls and very few boys. Boys and girls who score between 0.40 and 0.60 likely practice a combination of gender-typical and gender-atypical behaviors, whereas those who score greater than 0.60 are primarily gender typical in their behaviors. We consider respondents to be gender atypical in their behavior if they score lower than 0.40 (relative to their peers). We include a squared term for gender typicality to account for nonlinearity in relationship between gender typicality and GPA.[6]

High school GPA. The GPA variables are drawn from students' high school transcripts. We use the sampling weights provided by Add Health to adjust for transcript or survey nonresponse to reduce bias in estimates and standard errors for three GPA variables: (1) overall GPA, (2) math GPA, and (3) English GPA. Overall GPA is a cumulative measure for all high school years and subjects. English and math GPAs reflect the average GPA across all high school years in which a respondent was enrolled in that subject, such that if a student took math for three years during high school, her math GPA captures the average of the grades received in those courses. All GPA variables are continuous ranging from 4 (highest) to 0 (lowest). Full transcript data for all high school years exists for the vast majority of respondents in wave I because even students who were in ninth grade in wave I completed high school by the time transcript data were collected in wave III.[7]

Race. Race is measured via dummy variables for white, black, Hispanic, and Asian, with whites serving as the reference group.

Parent's education. We use parents' highest education as a proxy for socioeconomic status. The educational attainment of respondents' resident mothers and fathers were coded into the following categories: not completed high school, earned a high school degree, earned some college credit, and earned a four-year college degree or higher. With these data, we created a variable for the highest level of education of either parent, with less than a high school degree as the reference category.[8]

Father present. A dummy variable indicates whether a father figure—either biological, step, foster, or adoptive father— resided in the respondent's home at the time of the survey.

Picture vocabulary test. The Add Health Picture Vocabulary Test (PVT) is a shorter version of the nationally standardized Peabody PVT. It measures verbal and scholastic abilities and aptitude predictive of cognitive ability (see Halpern et al. 2000). Scores range from 14 to 137 and the measure indicates the quartile of the distribution in which the respondent scored.

Table 2 reports summary statistics for all variables. Boys' and girls' average gender typicality scores are 0.59 and 0.67, respectively, and significantly differ—a fact we discuss below. Consistent with prior research (DiPrete and Buchmann 2013), we find that, on average, boys earn significantly lower scores than girls across all three GPA measures; boys' and girls' mean demographic and family statistics are similar, as are mean PVT scores within each quartile.

Analysis Plan

We report gender typicality score distributions and use logistic regression for boys and girls separately to assess the characteristics associated with different gender typicality scores. Then we examine how gender typicality is associated with GPA for boys and for girls. Because GPAs within schools are likely correlated with and due to exposure to the same curriculum, scoring systems, or similar socioeconomic status of students, we use multilevel regression models to account for GPA variance across schools. Then, in order to examine whether the patterns vary by race/ethnicity, and socioeconomic status, we test for interactions between gender typicality and these variables on boys' and girls' GPAs.

Results

Distribution of Gender Typicality among Girls and Boys

Figure 1 illustrates the percentage distributions for each sex across the gender typicality scale. Notably, gender typicality peaks between 0.80 and 0.90 for girls and between 0.60 and 0.70 for boys, suggesting that girls tend to practice a narrower set of gendered behaviors based on this scale. Note that girls' higher typicality peak also stems from the fact that so few boys report practicing multiple feminine-typed behaviors. Indeed, given that gender typicality is a measure of the predictability of being a boy or girl, girls' higher score in gender typicality reflects the strong probability of being a girl when an individual reports multiple feminine practices.

Table 2: Variables used in analysis for boys and girls.

	Boys Percentage Distribution / Mean (SE)	Girls Percentage Distribution / Mean (SE)
Gender Typicality	0.59	0.67
	(0.01)	(0.01)
Transcript Grades		
Overall GPA	2.47	2.76
	(0.05)	(0.04)
English GPA	2.22	2.64
	(0.06)	(0.05)
Math GPA	2.13	2.37
	(0.04)	(0.04)
Race		
White	0.72	0.75
Black	0.17	0.17
Asian	0.04	0.02
Hispanic	0.07	0.06
Parent's Education		
Less than High School	0.10	0.11
High School	0.28	0.31
Some College	0.31	0.28
Bachelors or Higher	0.31	0.30
Father Present in Home	0.78	0.77
Picture Vocabulary Test		
Lowest Quartile	84.27	85.38
	(0.73)	(0.46)
Lower-mid Quartile	96.98	96.92
	(0.12)	(0.16)
Upper-mid Quartile	107.96	107.71
	(0.28)	(0.18)
Highest Quartile	119.11	118.62
	(0.46)	(0.29)
N	3,455	3,769

Notes: Sample is based on wave I Add Health In-home and Transcript Survey data. Statistics are based on weighted population data. Picture Vocabulary Test scores range from 14 to 137, with the values representing the average score for each quartile. GPA, grade point average.

For example, the item that measures the frequency of crying has the highest coefficient ($b = -1.47$) of all scale items, suggesting that a respondent who reports crying frequently is likely to be a girl because few boys report that they cry frequently.

We find no differences in the distribution of gender typicality by socioeconomic status as measured by parental education, but some noteworthy differences by race and ethnicity. Figure 2 presents a kernel density plot that demonstrates race-specific frequency distributions of gender typicality scores for boys and girls. Kernel density estimators generate the density f(x) from observations on x. The distribution is unimodal for boys and shows that black boys exhibit the most gender-typical

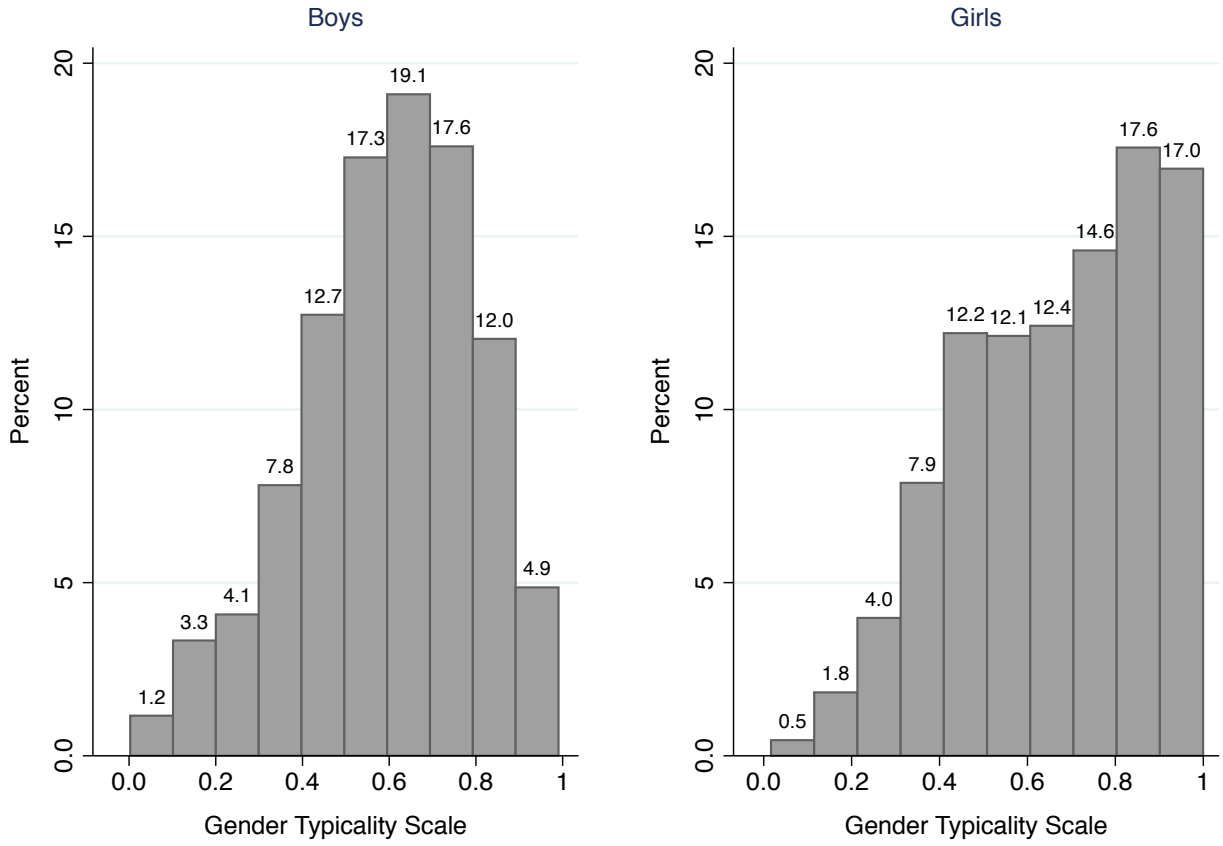

Figure 1: Percentage of high school students across the gender typicality scale.

behaviors. Differences are most notable between black and Asian boys. Based on numerical counts (not shown), about 58 percent of black boys score in the gender-typical range above 0.60—in contrast to the scores of Asian boys of which 46 percent are in this range. Nearly one-quarter of Asian boys have gender-atypical scores lower than 0.40, but only 13 percent of black and 16 percent of white boys score in this range. The percentages of white and Hispanic boys who score in the gender-typical range (53 percent and 50 percent, respectively) fall in between those of black and Asian boys.

Figure 2 shows a bimodal distribution for girls, in which gender typicality scores peak around 0.45 to 0.55 and again around 0.80 to 0.90. Notably, at the higher peak, the distribution of white, Asian, and Hispanic girls is denser than the distribution of black girls. Black girls have a greater density of observations at the lower peak, around the midpoint, suggesting that the largest proportion of black girls practice gender-balanced behaviors. In sum, these differences suggest that black girls and Asian boys practice less rigid gendered behaviors, as captured by the current scale, than their counterparts from other racial groups. Next, we examine whether these racial differences in gender typicality hold with more rigorous testing.

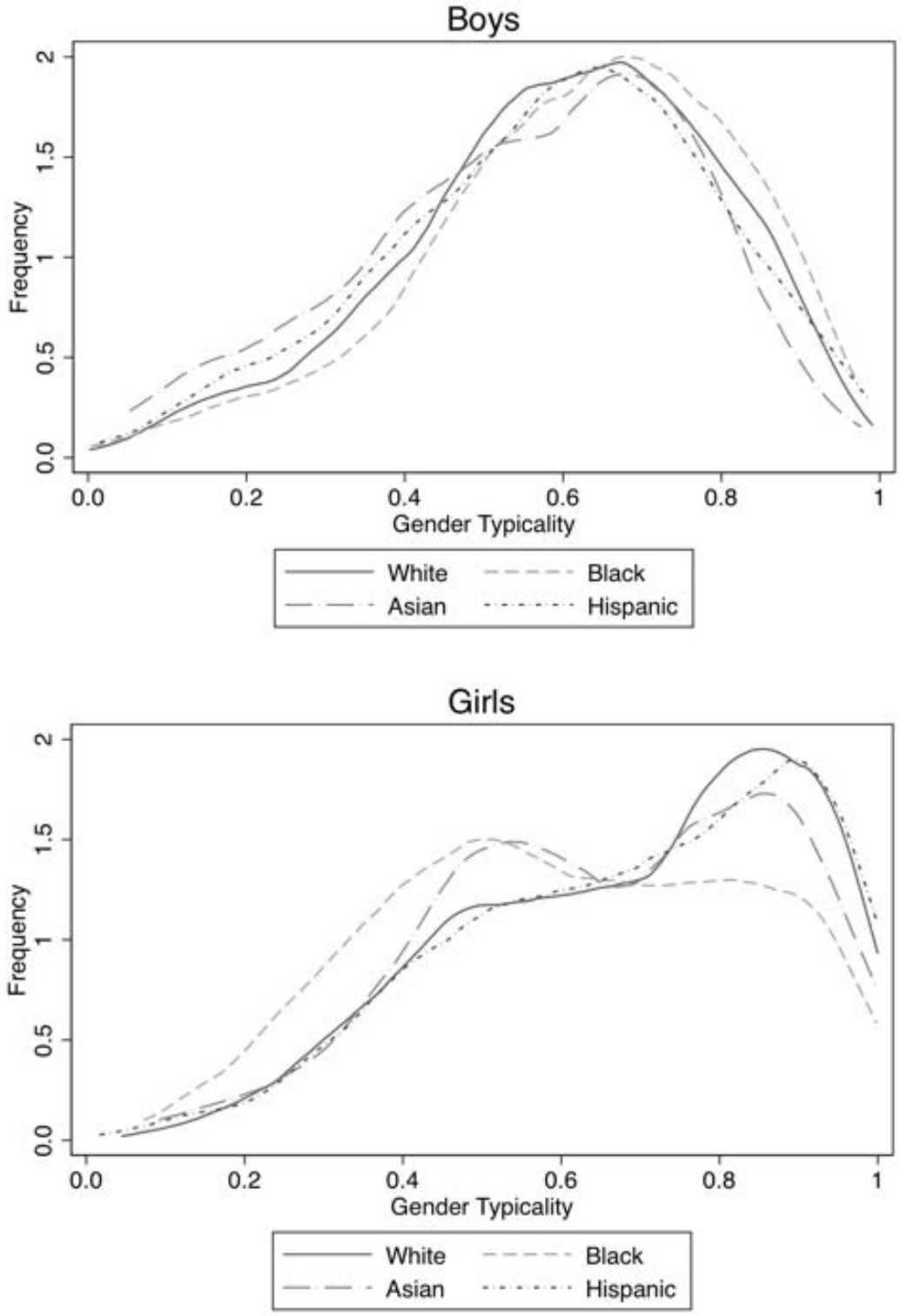

Figure 2: Gender typicality by race.

Table 3: Regression results that show relationship between respondent characteristics and gender typicality among boys and girls.

	Boys (1)	Girls (2)
Race		
White	–	–
Black	0.03*	−0.08†
	(0.02)	(0.02)
Asian	−0.03	0.01
	(0.02)	(0.02)
Hispanic	0.00	0.01
	(0.03)	(0.02)
Parent's Highest Education Level		
Less than high school degree	–	–
High school degree	0.02	0.02
	(0.02)	(0.03)
Some college credit	0.02	0.00
	(0.02)	(0.03)
College or advanced degree	0.02	−0.02
	(0.02)	(0.03)
Constant	0.54†	0.70†
	(0.03)	(0.04)
Observations (N)	3,455	3,769

Notes: Robust standard errors in parentheses. Model contains all control variables. See online supplement Table A1 for full model. $^{*}p < 0.05$, $^{\dagger}p < 0.01$.

Characteristics Associated with Gender Typicality

The regression analysis shown in Table 3 confirms the descriptive differences in gender typicality by race. Net of controls for parent's education, family structure, and PVT scores, the gender-typicality coefficients for white and black boys are significantly different, whereas the gender-typicality coefficients for Hispanic and Asian boys do not significantly differ from that of white boys. Black girls score significantly lower on the gender-typicality scale relative to white girls. By varying the racial reference group, we find that Asian boys are significantly less gender typical than black boys and that Asian and Hispanic girls are significantly more gender typical than black girls, net of controls (not shown). Notably, socioeconomic status, as measured by parent's education, is not significantly associated with gender typicality. Table A1 in the online supplement reports the full model and shows that boys in the lower quartile of PVT scores are also significantly higher on gender typicality relative to boys who score in the highest quartile of PVT scores. There is no significant association between PVT scores and gender typicality for girls.

Associations between Gender Typicality and Academic Performance

Next we test the relationship between gender typicality and overall and subject-specific GPAs, net of controls. Recall that gender typicality is defined as scoring greater than 0.60 and gender atypicality as scoring less than 0.40. Figure 3 presents predicted values from the multilevel regression models in graphical form and Tables A2 (boys) and A3 (girls) of the online supplement show the full regression models.

As illustrated in Figure 3, there is a significant nonlinear relationship between gender typicality and overall GPA for boys, net of controls. Boys' overall GPAs increase at a decelerating rate ($b_{GenderTypicality}$ = 0.88, $p < .01$) at very low gender typicality scores (i.e., when boys are very gender atypical) until about 0.37 (= –0.88/(2*−1.20); y = a + 0.88*X − 1.20*Xsq + e) on the scale, at which point the relationship reverses and becomes negative. Beyond the gender typicality score of 0.37, GPA significantly decreases at an accelerating rate ($b_{GenderTypicalitySquared}$ = −1.20, $p < 0.001$) and reaches its lowest level among boys who have the highest gender-typicality scores. Thus, boys who practice gender-atypical behaviors achieve the highest overall GPAs and those who practice gender-typical behaviors attain the lowest overall GPAs, net of controls.

Consistent with prior literature (Kao and Thompson 2003; Hsin 2018), Asian boys earn higher overall high school GPAs (b_{Asian} = 0.25, $p < 0.05$) than white boys and black boys; b_{Asian} = 0.27, $p < 0.10$ (in model where black is the reference category, not shown), even after controlling for parent's education, family structure, and PVT scores (see Table A2 of the online supplement). Notably, if boys earn the highest overall GPAs when they practice moderate gender atypicality, racial differences in gender atypicality be related to racial differences in overall GPAs. Recall that about a quarter of Asian boys but only 13 percent of black boys score in the gender-atypical range (less than 0.40).

Similar to the findings for overall GPA, a slight nonlinear pattern emerges between gender typicality and boys' English GPAs ($b_{GenderTypicality}$ = 0.72, p is approaching significance at 0.12; $b_{GenderTypicalitySquared}$ = –1.14, $p < 0.01$). This suggests a positive relationship between gender typicality and boys' English GPAs up until a turning point of 0.32 (= −0.72/(2*−1.14); y = a + 0.72*X − 1.14*Xsq + e). Indeed, Figure 3 shows that boys who practice moderate levels of gender-atypical behaviors outperform other boys, whereas boys with the highest gender-typicality scores earn the lowest English GPAs. In contrast to the slightly nonlinear relationship between gender typicality and these GPA measures, there is a strong negative linear relationship between gender typicality and boys' math GPAs ($b_{GenderTypicality}$ = −0.52, $p < 0.001$), suggesting that increases in boys' gender typicality are associated with lower math GPAs.

For girls, Figure 3 shows a significant nonlinear relationship between gender typicality and overall GPA (see also Table A3 of the online supplement). Specifically, girls' overall GPAs significantly increase at a decelerating rate ($b_{GenderTypicality}$ = 1.01 $p < 0.01$) until about 0.65 (= −1.01/(2*−0.78); y = a + 1.01*X − 0.78*Xsq + e) on the gender-typicality scale, at which point the relationship reverses and becomes negative. Beyond typicality scores of 0.65, rising typicality scores are significantly associated with greater GPA declines ($b_{GenderTypicalitySquared}$ = −0.78, $p < 0.05$). Thus, girls who practice multifaceted behaviors (or slightly gender-typical behaviors)

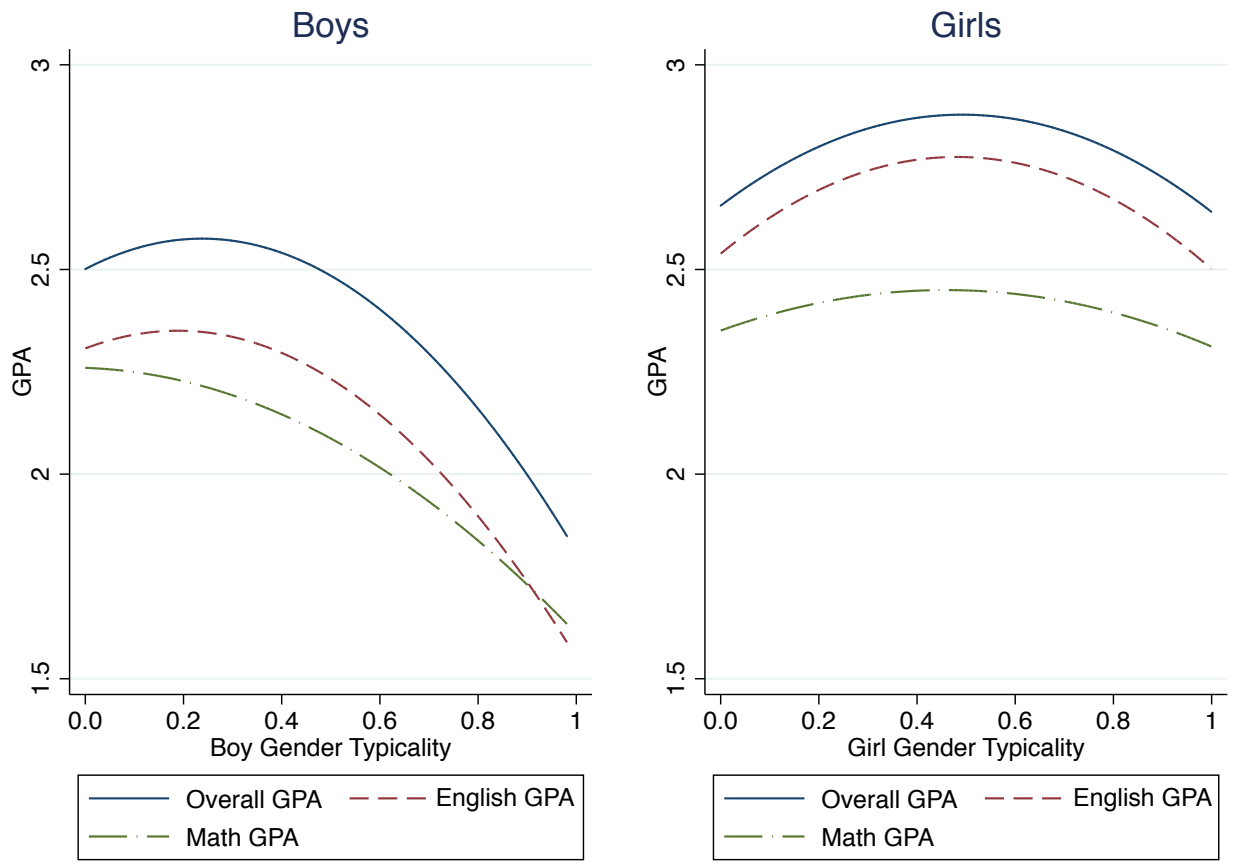

Figure 3: Boys' and girls' predicted transcript grade point averages based on gender typicality.

earn the highest overall GPA relative to girls at the tails of the distribution who display either very gender-typical or very gender-atypical behaviors.

There is also a significant nonlinear relationship between gender typicality and girls' English GPAs ($b_{GenderTypicality} = 1.12$, $p < 0.05$; $b_{GenderTypicalitySquared} = -0.88$, $p < 0.05$), with girls who practice multifaceted and slightly gender-typical behaviors achieving the highest English GPAs, net of race, parents' education, family structure, and PVT scores. For girls, we find no relationship between gender typicality and math GPA; neither gender typicality nor the squared term is significantly associated with girls' math GPAs.

As is clear from Figure 3, girls exhibit far less variance in their GPAs across the gender-typicality scale relative to boys, suggesting that boys' GPAs are more strongly related to their gendered behavioral orientations. For example, there are sizeable GPA gaps between moderately gender-typical boys and moderately gender-atypical boys; for girls, there are no such gaps. For example, boys with a gender typicality squared value of less than 0.25 have a 2.56 overall GPA, on average, whereas boys with a gender typicality squared value of greater than 0.75 have about a 2.10 overall GPA, on average—a difference of nearly half of a grade

point. In contrast, girls who score less than 0.25 or greater than 0.75 on gender typicality both have average GPAs around 2.70.

Differences by Race and Socioeconomic Status

Figure 2 shows racial differences in the distribution of gender typicality, but based on models that include interactions between gender typicality and race (not shown), we find that the relationship between gender typicality and academic performance largely operates similarly for black, Hispanic, Asian, and white adolescents.

We find few significant interaction terms between gender typicality and parents' education for boys; however, these interaction terms are significant for all three girls' GPA measures. Thus, here we concentrate on results pertaining to girls (see Table A4 of the online supplement). Figure 4 shows how the relationship between overall GPA and gender typicality differs according to girls' socioeconomic status as measured by the highest education level of either parent. Comparing girls with a parent who did not complete high school to those with a parent who has at least a college degree, GPA disparities are magnified when girls practice multifaceted gender behaviors (between 0.4 and 0.6 on the typicality scale) and moderate levels of gender typicality (between 0.6 and 0.8 on the typicality scale). However, at the highest levels of gender typicality, GPA disparities between girls of different socioeconomic statuses are smaller. These results suggest that the relationship between extreme gender typicality and GPA is more similar among girls of different socioeconomic statuses. This is an important finding considering that about 35 percent of girls score 0.80 or higher and about 17 percent score 0.90 or higher on the gender-typicality scale.

Discussion

This study is the first to use nationally representative data to examine how gender typicality is related to boys' and girls' high school GPAs. The Add Health data are ideal for testing the extent to which gender typicality is related to U.S. boys' and girls' educational achievement in high school as it provides a representative sample of U.S. schools in terms of school size, racial composition, type, and region. The findings suggest that the degree to which students, and especially boys, practice behaviors typical of their gender is strongly related to their academic performance, even net of other well-established characteristics such as race, socioeconomic background, and cognitive ability.

This study advances the literature in several ways. It reveals that high school boys who report moderate levels of gender atypicality (around 0.37) earn the highest overall GPAs, net of controls, but very few boys score in this range. As boys' gender-normative behaviors increase, their average overall and English GPAs decline steeply. These results substantiate those of qualitative studies, suggesting that masculinity practices are detrimental to boys' academic achievement. These findings also challenge popular claims that the weakening of masculinity and boys' identity is a source of some boys' academic underperformance relative to their abilities (Tyre 2009; Sommers 2013). In fact, we find that gender atypicality

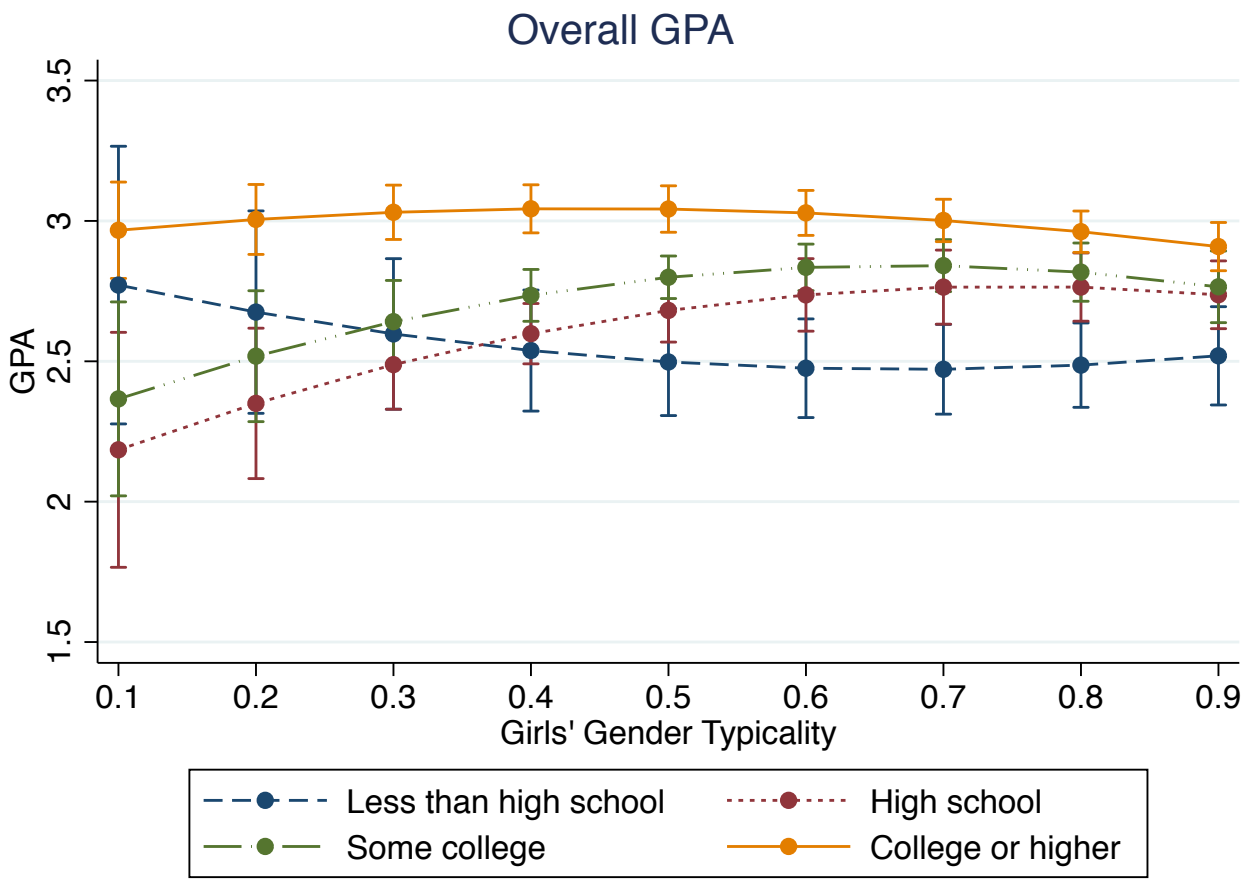

Figure 4: Adjusted predictions of girls' grade point averages by parents' education levels and gender typicality.

among boys is associated with higher academic performance. Also striking is the strong negative linear relationship between gender typicality and boys' math GPA, given that mathematics is a masculine-typed subject (Leaper, Farkas, and Brown 2012), especially in the 1990s when these data were collected. It is possible that, as awareness of girls' strides in the academic realm has grown in recent decades, more boys may have come to see striving in school as feminine regardless of the subject and that their academic underperformance has become more prevalent since these data were collected.

The findings also have important implications for racial-gender differences in academic performance. Specifically, based on the scale used, black boys appear to adhere more to some gender-typical behaviors than white boys, and black girls adhere less to some gender-typical behaviors than white girls. Practicing moderate gender typicality may protect black girls from other race-based disadvantages that contribute to otherwise lower GPAs of black girls relative to white girls. Black boys' greater tendency to practice some gender-typical behaviors, particularly compared to white and Asian boys, may be related to their lower average GPAs relative to other groups (Kao and Thompson 2003). Although we find that gender typicality is similarly related to GPA for all race and ethnic groups, race-based

differences in gender typicality may be related to the size of the gender gap in academic achievement within racial groups, such as the particularly larger gender gap in academic achievement and educational attainment among black adolescents (McDaniel et al. 2011).

This research also contributes new knowledge about the relationship between gender typicality and girls' GPA. Their overall and English GPAs peak when girls practice slight gender typicality (instead of gender atypicality as is the case for boys). Although few boys score in the gender-atypical range associated with peak academic performance, many girls score in the slight gender-typical range associated with this GPA apex. This finding could help explain why girls, on average, outperform boys in GPA, as more girls are positioned to potentially accrue advantages associated with practicing these gender norms. In this sense, girls' high performance does not appear to come at the expense of culturally valued gender presentations.

This study uncovers novel associations between gender typicality and academic performance, but it is not without limitations. First, the gender-typicality scale used here cannot capture the full range of gendered behaviors that may be related to academic achievement. We urge scholars to explore other gender-typicality scales in future research. Other scales that include different behaviors may yield different distributions in gender typicality for boys and girls. Second, additional research is required to evaluate how these patterns have changed in the more than 20 years since the Add Health data were collected. As noted above, it is possible that boys' underperformance has become more prevalent with growing awareness of girls strides in the academic realm, but it is also possible that it has become less prevalent. Some research indicates growing societal acceptance of adolescents who do not conform to normative gender roles as well as a decoupling of some traits from either femininity or masculinity. Attending to one's appearance does not necessarily indicate femininity (McCormack 2010), and being athletic is no longer viewed as a masculine trait (Quatman, Sokolik, and Smith 2000). At the same time, feminist scholars contend that much of the expansion of gender norms has been bestowed on girls, not boys (England 2010; Coontz 2013). Girls have the ability to practice some normative masculine behaviors without compromising their gender status—in fact, practicing nonnormative gender behaviors may serve to boost girls' social status among peers (Pascoe 2007). Thus, to the degree that normative femininities have expanded to incorporate some traditionally masculine traits, our research may underestimate the associations between girls' multifaceted gendered behaviors and their academic achievement. More recent survey data that can be linked to academic transcript data are necessary to unpack the relationships between gendered behaviors and academic outcomes in the current era.

As long as gender-atypical boys experience more teasing and are rated as less popular than gender-typical boys (Pascoe 2007; Jewell and Brown 2014), some boys may continue to engage in gender-typical behaviors that undermine their academic achievement to preserve their social status. Importantly, our research counters popular understandings of the reasons behind some boys' poor school performance and outcomes. As Coontz (2013) points out, "Many experts today blame men's troubles on the weakening of their traditional gender identity. Boys, they complain,

are being forced to 'act like girls' in school." This study shows that some boys' poor academic performance may be related to the pressure they feel to "act like a boy" and that the persistence of a gender system that values boys' masculinity practices derails their academic progress. Thus, the continued broadening of gender norms that has been advantageous for girls would be advantageous for all students.

Notes

1 Add Health is a school-based, longitudinal study of health-related behaviors of 7th to 12th graders and their subsequent outcomes in young adulthood. Using a clustered sampling design, Add Health first sampled schools and then sampled students within the schools, resulting in a sample of 18,924 students in 52 middle schools and 80 high schools who completed an in-school survey during the 1994 to 1995 school year. Incorporating systematic sampling methods and implicit stratification into the Add Health study design ensured this sample is representative of U.S. schools with respect to region of the country, urbanicity, school size, school type, and ethnicity.

2 For a detailed description of the methods used in the AHAA, see http://www.laits.utexas.edu/ahaa/home.

3 Although Add Health attempted to collect transcript release forms from all students, transcripts were missing for some students for the following reasons: the student did not agree to participate, did not attend high school, was home-schooled, attended high school outside of the United States, or because the school had closed or provided incomplete or erroneous transcripts (Riegle-Crumb et al. 2005). The AHAA study computed and provided sample weights to users in order to reduce the bias due to missing transcripts. In analyses not shown, we find that respondents who provided transcript data did not significantly differ in terms of gender typicality from respondents who did not release their transcript data.

4 There are not straightforward guidelines for multilevel models that have covariate variables with missing values. Grund, Ludtke, and Robitzsch (2016) evaluated the suitability of different techniques for handling missing data for multilevel models and found that listwise deletion in random slope models (such as ours) provided "practically unbiased estimates of the slope variance if the sample size was sufficiently large" (P. 645). As such and given the small amount of missing data for our covariates, we use listwise deletion. The missing values dropped were as follows: 79 for father's presence in the home, 72 for race, 93 for gender typicality (where a participant did not answer one of the 12 items in the scale), 396 for PVT scores, 278 for region, and 283 for sampling weights. We also dropped 305 Native Americans because of their small sample size.

5 The questions that appear in the scale used by Cleveland et al. (2001) but are not included in our scale because they were not asked in the wave I survey are: Do you live without thought for future? How sensitive are you to others' feelings? Do you like to take risks? How emotional are you?

6 Notably, Add Health did not distinguish transgender respondents from cisgender respondents, as different scales to identify gender identity are only now becoming more common in surveys (Westbrook and Saperstein 2015). In this study, we use the term male to correspond with the term boy and female to correspond with the term girl, even though we recognize that not all respondents who answered male or female necessarily identified as a boy or girl, respectively.

7 Nearly 90 percent of sample respondents had four or more years of high school transcript data and about 95 percent had at least three or more years of high school transcript data for the GPA measures.

8 We used a scaffold approach to obtain accurate estimates of mothers' and fathers' educational levels. For example, if a mother (father) had filled out her (his) education during the in-home wave I survey, we used that value; if not, then we used the father's (mother's) response of mother's (father's) education in the in-home wave I survey. If both values were missing, we used the child's in-home or in-school survey report of mother's (father's) education. This approach led to no missing values for this measure.

References

Archer, Louise, Jennifer DeWitt, Jonathan Osborne, Justin Dillon, Beatrice Willis, and Billy Wong. 2012. "'Balancing Acts': Elementary School Girls' Negotiations of Femininity, Achievement, and Science." *Science Education* 96:967–89. https://doi.org/10.1002/sce.21031.

Archer, Louise, Anna Halsall, and Sumi Hollingworth. 2007. "Class, Gender, (Hetero) Sexuality and Schooling: Paradoxes within Working-Class Girls' Engagement with Education and Post-16 Aspirations." *British Journal of Sociology of Education* 28:165–80. https://doi.org/10.1080/01425690701192570.

Auster, Carol J., and Susan C. Ohm. 2000. "Masculinity and Femininity in Contemporary American Society: A Reevaluation Using the Bem Sex-Role Inventory." *Sex Roles* 43:499–528. https://doi.org/10.1023/a:1007119516728.

Bem, Sandra L. 1974. "The Measurement of Psychological Androgyny." *Journal of Consulting and Clinical Psychology* 42:155–62.

Bowen, William G., Matthew M. Chingos, and Michael S. McPherson. 2009. *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton, NJ: Princeton University Press.

Brown, Lyn Mikel. 1997. "Performing Femininities: Listening to White Working-Class Girls in Rural Maine." *Journal of Social Issues* 53:683–701. https://doi.org/10.1111/0022-4537.00044.

Buchmann, Claudia, and Thomas A. DiPrete. 2006. "The Growing Female Advantage in College Completion: The Role of Family Background and Academic Achievement." *American Sociological Review* 71:515–41. https://doi.org/10.1177/000312240607100401.

Chu, Judy Y., Michelle V. Porche, and Deborah L. Tolman. 2005. "The Adolescent Masculinity Ideology in Relationships Scale: Development and Validation of a New Measure for Boys." *Men and Masculinities* 8:93–115. https://doi.org/10.1177/1097184X03257453.

Cleveland, H. Harrington, J. Richard Udry, and Kim Chantala. 2001. "Environmental and Genetic Influences on Sex-Typed Behaviors and Attitudes of Male and Female Adolescents." *Personality and Social Psychology Bulletin* 27:1587–98. https://doi.org/10.1177/01461672012712003.

Collins, Patricia Hill. 2000. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York, NY: Routledge.

Connell, R. W., and James W. Messerschmidt. 2005. "Hegemonic Masculinity: Rethinking the Concept." *Gender and Society* 19:829–59. https://doi.org/10.1177/0891243205278639.

Connell, R. W. 2005. *Masculinities*. 2nd ed. Berkeley, CA: University of California Press.

Coontz, Stephanie. 2013. "Yes, I've Folded up My Masculine Mystique, Honey." *The Sunday Times*, February 24. Retrieved August 30, 2019, http://www.thesundaytimes.co.uk/sto/newsreview/article1219753.ece.

Davis, Shannon N., and Theodore N. Greenstein. 2009. "Gender Ideology: Components, Predictors, and Consequences." *Annual Review of Sociology* 35:87–105. https://doi.org/10.1146/annurev-soc-070308-115920.

DiPrete, Thomas A., and Claudia Buchmann. 2013. *The Rise of Women: The Growing Gender Gap in Education and What It Means for American Schools*. New York, NY: Russell Sage Foundation.

England, Paula. 2010. "The Gender Revolution Uneven and Stalled." *Gender and Society* 24:149–66. https://doi.org/10.1177/0891243210361475.

Ferguson, Ann Arnett, Martha Minow, Elaine Scarry, and Austin Sarat. 2000. *Bad Boys: Public Schools in the Making of Black Masculinity*. Ann Arbor, MI: University of Michigan Press.

Graham, John W., Allison E. Olchowski, and Tamika D. Gilreath. 2007. "How Many Imputations Are Really Needed? Some Practical Clarifications of Multiple Imputation Theory." *Prevention Science* 8:206–13. https://doi.org/10.1007/s11121-007-0070-9.

Grund, Simon, Oliver Lüdtke, and Alexander Robitzsch. 2016. "Multiple Imputation of Missing Covariate Values in Multilevel Models with Random Slopes: A Cautionary Note." *Behavior Research Methods* 48:640–49. https://doi.org/10.3758/s13428-015-0590-3.

Halpern, Carolyn Tucker, Kara Joyner, J. Richard Udry, and Chirayath Suchindran. 2000. "Smart Teens Don't Have Sex (or Kiss Much Either)." *Journal of Adolescent Health* 26:213–25. https://doi.org/10.1016/s1054-139x(99)00061-0.

Heyder, Anke, and Ursula Kessels. 2013. "Is School Feminine? Implicit Gender Stereotyping of School as a Predictor of Academic Achievement." *Sex Roles* 69:605–17. https://doi.org/10.1007/s11199-013-0309-9.

Hsin, Amy. 2018. "Hegemonic Gender Norms and the Gender Gap in Achievement: The Case of Asian Americans." *Sociological Science* 5:752–74. https://doi.org/10.15195/v5.a32.

Jackson, Carolyn, and Steven Dempster. 2009. "'I Sat Back on My Computer ... with a Bottle of Whisky next to Me': Constructing 'Cool' Masculinity through 'Effortless' Achievement in Secondary and Higher Education." *Journal of Gender Studies* 18:341–56. https://doi.org/10.1080/09589230903260019.

Jewell, Jennifer A., and Christia Spears Brown. 2014. "Relations among Gender Typicality, Peer Relations, and Mental Health During Early Adolescence." *Social Development* 23:137–56. https://doi.org/10.1111/sode.12042.

Kao, Grace, and Jennifer S. Thompson. 2003. "Racial and Ethnic Stratification in Educational Achievement and Attainment." *Annual Review of Sociology* 29:417–42. https://doi.org/10.1146/annurev.soc.29.010202.100019.

Kreiger, Tyson C., and Larry E. Dumka. 2006. "The Relationships between Hypergender, Gender, and Psychological Adjustment." *Sex Roles* 54:777–85. https://doi.org/10.1007/s11199-006-9044-9.

Leaper, Campbell, Timea Farkas, and Christia Spears Brown. 2012. "Adolescent Girls' Experiences and Gender-Related Beliefs in Relation to Their Motivation in Math/Science and English." *Journal of Youth and Adolescence* 41:268–82. https://doi.org/10.1007/s10964-011-9693-z.

Legewie, Joscha, and Thomas A. DiPrete. 2012. "School Context and the Gender Gap in Educational Achievement." *American Sociological Review* 77:463–85. https://doi.org/10.1177/0003122412440802.

Lippa, Richard A. 1998. "Gender-Related Individual Differences and National Merit Test Performance: 'Feminine' Boys and 'Masculine' Girls Tend to Do Better." Pp. 177–94 in *Males, Females, and Behavior: Toward Biological Understanding*, edited by Lee Ellis and Linda Ebertz. Westport, CT: Praeger Publishers.

Lippa, Richard A. 2000. "Gender-Related Traits in Gay Men, Lesbian Women, and Heterosexual Men and Women: The Virtual Identity of Homosexual-Heterosexual Diagnosticity and Gender Diagnosticity." *Journal of Personality* 68:899–926. https://doi.org/10.1111/1467-6494.00120.

Lippa, Richard A., and Sharon Connelly. 1990. "Gender Diagnosticity: A New Bayesian Approach to Gender-Related Individual Differences." *Journal of Personality and Social Psychology* 59:1051–65. https://doi.org/10.1037/0022-3514.59.5.1051.

Livingston, Robert W., Ashleigh Shelby Rosette, and Ella F. Washington. 2012. "Can an Agentic Black Woman Get Ahead? The Impact of Race and Interpersonal Dominance on Perceptions of Female Leaders." *Psychological Science* 23:354–58. https://doi.org/10.1177/0956797611428079.

MacLeod, Jay. 2008. *Ain't No Makin' It: Aspirations and Attainment in a Low-Income Neighborhood*. 3rd ed. Boulder, CO: Westview Press.

McCormack, Mark. 2010. "Changing Masculinities in Youth Cultures." *Qualitative Sociology* 33:111–15. https://doi.org/10.1007/s11133-009-9146-0.

McDaniel, Anne, Thomas A. DiPrete, Claudia Buchmann, and Uri Shwed. 2011. "The Black Gender Gap in Educational Attainment: Historical Trends and Racial Comparisons." *Demography* 48:889–914. https://doi.org/10.1007/s13524-011-0037-0.

Morris, Edward W. 2007. "'Ladies' or 'Loudies'? Perceptions and Experiences of Black Girls in Classrooms." *Youth and Society* 38:490–515. https://doi.org/10.1177/0044118x06296778.

Morris, Edward W. 2008. "'Rednecks,' 'Rutters,' and 'Rithmetic: Social Class, Masculinity, and Schooling in a Rural Context." *Gender and Society* 22:728–51. https://doi.org/10.1177/0891243208325163.

Morris, Edward W. 2012. *Learning the Hard Way: Masculinity, Place, and the Gender Gap in Education*. New Brunswick, NJ: Rutgers University Press.

Pascoe, C. J. 2007. *Dude, You're a Fag: Masculinity and Sexuality in High School*. 1st ed. Oakland, CA: University of California Press.

Pyke, Karen D. 1996. "Class-Based Masculinities The Interdependence of Gender, Class, and Interpersonal Power." *Gender and Society* 10:527–49. https://doi.org/10.1177/089124396010005003.

Qin, Desiree Baolin. 2009. "Being 'Good' or Being 'Popular': Gender and Ethnic Identity Negotiations of Chinese Immigrant Adolescents." *Journal of Adolescent Research* 24:37–66. https://doi.org/10.1177/0743558408326912.

Quatman, Teri, Elizabeth Sokolik, and Karin Smith. 2000. "Adolescent Perception of Peer Success: A Gendered Perspective over Time." *Sex Roles* 43:61–84. https://doi.org/10.1023/a:1007039712348.

Riegle-Crumb, Catherine, Chandra Muller, Kenneth Frank, and Kathryn S. Schiller. 2005. "National Longitudinal Study of Adolescent Health: Wave III Education Data Weights Code Book." Carolina Population Center, UNC. http://www.laits.utexas.edu/ahaa/docs/Weights_2005_Riegle-Crumb.pdf.

Roderick, Melissa, Vanessa Coca, and Jenny Nagaoka. 2011. "Potholes on the Road to College: High School Effects in Shaping Urban Students' Participation in College Application,

Four-Year College Enrollment, and College Match." *Sociology of Education* 84:178–211. https://doi.org/10.1177/0038040711411280.

Rubin, Donald B. 1996. "Multiple Imputation After 18+ Years." *Journal of the American Statistical Association* 91:473–89. https://doi.org/10.2307/2291635.

Schrock, Douglas, and Michael Schwalbe. 2009. "Men, Masculinity, and Manhood Acts." *Annual Review of Sociology* 35:277–95. https://doi.org/10.1146/annurev-soc-070308-115933.

Seem, Susan Rachael, and M. Diane Clark. 2006. "Healthy Women, Healthy Men, and Healthy Adults: An Evaluation of Gender Role Stereotypes in the Twenty-First Century." *Sex Roles* 55:247–58. https://doi.org/10.1007/s11199-006-9077-0.

Simmons, Roberta G., Richard Burgeson, Steven Carlton-Ford, and Dale A. Blyth. 1987. "The Impact of Cumulative Change in Early Adolescence." *Child Development* 58:1220–34. https://doi.org/10.1111/j.1467-8624.1987.tb01453.x.

Smiler, Andrew P., and Marina Epstein. 2010. "Measuring Gender: Options and Issues." Pp. 133–57 in *Handbook of Gender Research in Psychology*, edited by Joan C. Chrisler and Donald R. McCreary. New York, NY: Springer.

Sommers, Christina Hoff. 2013. *The War Against Boys: How Misguided Policies Are Harming Our Young Men*. New York, NY: Simon and Schuster.

Terman, Lewis, and Catherine Cox Miles. 1936. *Sex and Personality*. New Haven, CT: Yale University Press.

Tyre, Peg. 2009. *The Trouble with Boys: A Surprising Report Card on Our Sons, Their Problems at School, and What Parents and Educators Must Do*. New York, NY: Harmony Books.

Udry, J. Richard, and Kim Chantala. 2004. "Masculinity-Femininity Guides Sexual Union Formation in Adolescents." *Personality and Social Psychology Bulletin* 30:44–55. https://doi.org/10.1177/0146167203258840.

Ueno, Koji, and Summer McWilliams. 2010. "Gender-Typed Behaviors and School Adjustment." *Sex Roles* 63:580–91. https://doi.org/10.1007/s11199-010-9839-6.

Ueno, Koji, Teresa Roach, and Abráham E. Peña-Talamantes. 2013. "Sexual Orientation and Gender Typicality of the Occupation in Young Adulthood." *Social Forces* 92:81–108. https://doi.org/10.1093/sf/sot067.

West, Candace, and Don H. Zimmerman. 1987. "Doing Gender." *Gender and Society* 1:125–51.

Westbrook, Laurel, and Aliya Saperstein. 2015. "New Categories Are Not Enough: Rethinking the Measurement of Sex and Gender in Social Surveys." *Gender and Society* 29:534–60. https://doi.org/10.1177/0891243215584758.

Willis, Paul E. 1977. *Learning to Labor: How Working Class Kids Get Working Class Jobs*. New York, NY: Columbia University Press.

Acknowledgments: This research uses data from the National Longitudinal Study of Adolescent Health (Add Health), a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis. This research also uses data from the Adolescent Health and Academic Achievement study, which was funded by a grant (R01 HD040428-02, Chandra Muller, PI) from the National Institute of Child Health and Human Development and a grant (REC-0126167, Chandra Muller, PI, and Pedro Reyes, Co-PI) from the National Science Foundation. This research was also supported by grant 5 R24 HD042849, Population Research Center, awarded to the Population Research Center at The University of Texas at Austin by the Eunice Kennedy Shriver National Institute of Health and Child Development. Opinions reflect those of the authors and do not necessarily reflect those of the granting agencies. We thank the reviewers and editor for their helpful comments. We also are grateful to Yue Qian, Paula England, Tom DiPrete, and participants in the Seminar Series at the Center for the Study of Wealth and Inequality at Columbia University and participants in the University of Michigan Department of Sociology Seminar Series.

Jill E. Yavorsky: Department of Sociology, University of North Carolina Charlotte.
E-mail: jyavorsk@uncc.edu.

Claudia Buchmann: Department of Sociology, The Ohio State University.
E-mail: buchmann.4@osu.edu.