


## Going to College With a Posse: How Having High School Peers on Campus Supports College Achievement

Kevin R. Binning   
 Lorraine R. Blatt  
 Susie Chen  
 Elizabeth Votruba-Drzal  
 University of Pittsburgh

*The social experience of transitioning to a 4-year university varies widely among students. Some attend with few or no prior contacts or acquaintances from their hometown; others attend with a large network of high school alumni. Using a sample (N = 43,240) of undergraduates spanning 7.5 years at a public university, we examine what factors predict high school peer prevalence (HSPP) on campus and whether HSPP predicts college achievement above and beyond such factors. Analyses found that HSPP was predicted by variables associated with societal privilege (e.g., being White, continuing generation). Above and beyond these variables, HSPP independently predicted higher grades in gateway STEM (science, technology, engineering, and mathematics) courses and, among first-generation college students, higher retention. The role of HSPP in fostering equity and inequity is discussed. A preprint of this article is available at <https://psyarxiv.com/xhpuc/>.*

**Keywords:** admissions, equity, first-generation college students, hierarchical linear modeling, higher education, peer interaction/friendship, privilege, psychology, regression analyses, retention, social belonging, social capital, socioeconomic status

STUDENTS, educators, and policy makers have long assumed that when students go to college with demographically similar peers, it bolsters their chances for success. This assumption has driven policy around fostering equity in education for decades, including efforts to create “posses” of localized students to attend college together (e.g., Posse Foundation, 2020) and efforts to create learning communities for first-year students (e.g., Tinto, 2003). However, in the absence of intentional policy efforts, the size of students’ social networks at matriculation varies widely: while many students attend college relatively alone, with no or few contacts from their hometown, others attend with a large social network of peers from their high school. In the present research, we draw from an administrative data set of over 43,000 undergraduates from a 4-year public research university to understand both the sources and the consequences of students’ preexisting high school connections. Specifically, we examine (a) the demographic predictors of high school peer prevalence (HSPP) at matriculation and (b) whether HSPP independently predicts academic outcomes, including first-year STEM (science, technology, engineering, and mathematics) performance and college retention.

We argue that addressing these questions will help highlight the role of high school peers in shaping college success and point toward new avenues for understanding and promoting equity in higher education. Given that most universities

maintain information on students’ high school background, HSPP may be an inexpensive means to support students, for example, by identifying students with solo status and facilitating networking opportunities for them. On a more conceptual level, understanding HSPP may yield insight into how societal forces create uneven social and academic benefits for students, and how these uneven benefits may create more benefits that then propagate and reinforce inequity over time.

### Positionality Statement

We approach the present research from the perspective of social, developmental, and educational psychology. Our interest in developing a measure of HSPP arose from the body of evidence suggesting that college students, like people in general, are often exquisitely attuned to signals of belonging in their environments (e.g., Baumeister & Leary, 1995; Walton et al., 2012). People respond to subtle belonging cues in their social environments—such as the ethnic composition of their small workgroups (e.g., Binning et al., 2020), the ethnic makeup of their university (e.g., Binning & Unzueta, 2013), and the perceived inclusivity of organizational norms and structures (e.g., Greenaway & Turetsky, 2020; Murphy et al., 2007)—in the service of a fundamental need to belong (Baumeister & Leary, 1995). Establishing a sense of belonging in college is crucial for helping students



reach their full academic potential (e.g., Chen et al., 2021; Strayhorn, 2012; Walton & Brady, 2019), and research has shown that solo status, defined as being the only representative of a particular social category, can have detrimental effects on learning (Sekaquaptewa & Thompson, 2002), memory (Lord & Saenz, 1985), and self-regulation (Johnson & Richeson, 2009). As such, our core thesis is that HSPP uniquely predicts student outcomes above and beyond the demographic variables associated with HSPP. Furthermore, we reasoned that while HSPP may be generally beneficial for students, it may be especially beneficial for students who need it most, such as those from demographic backgrounds that typically have lower levels of belonging and less access to valuable information (e.g., first-generation college students).

Below we outline our rationale for understanding both the predictors of HSPP and why HSPP may independently predict college achievement.

### *Predictors of HSPP*

We hypothesize that having a large cohort of high school peers on campus at matriculation is likely to be more common among students from more privileged, historically overrepresented societal categories, and higher socioeconomic status (SES) backgrounds. A variety of evidence suggests that people from high social status backgrounds enjoy greater access to valued social networks (see Piff et al., 2018), which we argue extends to include networks of high school peers when attending college. Thus, while HSPP will tend to be higher among those with higher status backgrounds, it may be most beneficial for students from lower status backgrounds.

Selective admissions standards often have the effect of disproportionately filtering out particular groups of students. That is, with notable exceptions (e.g., historically Black colleges and universities), 4-year universities have long had an overrepresentation of privileged groups (e.g., White, high SES, males) relative to their proportion of the broader society. High schools, by contrast, serve a comparatively diverse and inclusive student population, but high schools are highly segregated in terms of factors like race/ethnicity and SES, and high-SES high schools tend to send far more students to college than do low-SES high schools (Rumberger & Palardy, 2005).

Such patterns logically imply that, all else being equal, students from more privileged backgrounds may be more likely to arrive on campus with the additional privilege of having a viable social network in place when they arrive. If a student attends college from a community in which only one in four students attend college, they will tend to have fewer peers on college campuses than a similar student from a community where 3 in 4 attend college. Thus, the high school to college filtering process,

and the inequities therein, may result in students from high-SES backgrounds being more likely to matriculate with preexisting social connections.

### *HSPP as a Predictor of College Achievement*

Why would having high school peers on campus at matriculation relate to later college achievement? Below we argue that HSPP is associated with belongingness cues and access to information, both of which may independently predict improved college outcomes.

*Peer Prevalence as a Source of Mere Belonging.* Several brain structures (e.g., dorsolateral prefrontal cortex) and processes are involved with social connection and belonging, as humans attend to belonging cues in their environment seemingly by default (for a review, see Lieberman, 2013). Together with well-known homophily biases, whereby people seek out others who share special background characteristics (McPherson et al., 2001), we reasoned that students likely attend to and encounter their high school peers at college at rates that are greater than would be predicted by their proportional representation on campus.

High school peers could be close friends or mere acquaintances. However, students may benefit from high school peers on campus even if they do not know them well. Having a connection to others in academic settings, even simple acquaintances, enhances academic motivation. Walton et al. have termed the superficial connections people share with others in performance domains “mere belonging” (Walton & Cohen, 2011; Walton et al., 2012). Distinct from social belonging, which connotes deeper feelings of acceptance and security, mere belonging describes minimal, chance, trivial, or superficial connections that people share with other people (Walton et al., 2012). These connections include things like sharing a birthday with another student, sharing the same taste in food as another student, and coming from the same high school as another student. In each of these cases, evidence suggests that mere belonging in a performance domain enhances performance motivation because it highlights socially shared performance goals (Walton et al., 2012). As such, mere belonging can enhance students’ sense of belonging, which in turn predicts both short- and long-term college success (Brady et al., 2020; Gopalan & Brady, 2020; Strayhorn, 2012; Walton & Cohen, 2011).

Furthermore, HSPP should be particularly helpful to new college students following the transition to college (Swenson et al., 2008). Tinto (1993) posited that experiences during this pivotal timepoint lead to academic and social integration, which then lead to college persistence. In other words, how students adjust to the college transition has critical long-term implications for their college success. Students who feel psychologically secure in their college environments may be more effective at cultivating their social networks (Turetsky

et al., 2020), which when connected to the present work suggests how “the rich get richer.” That is, HSPP may be one example of how societal privilege creates social circumstances that facilitate thriving. Students with social networks in place when they matriculate do not need to integrate into entirely new social networks; they arrive with one already in place (Hale et al., 2005; Hoffman et al., 2002). The head start provided by high HSPP may help students achieve the academic and social integration needed to maintain engagement and persistence in college over time.

*Peer Prevalence and Access to Information.* We argue that having many peers from high school elicits both bonding and bridging social capital (Putnam, 2000). Bonding social capital is similar to the idea of mere belonging: having high school peers on campus creates a means to bond with other students, as it denotes a shared high school and regional identity.

Importantly, high school peers also provide bridging capital, as the weak ties (Gee et al., 2017; Granovetter, 1973) and connections allow students better access to external features in the environment, such as access to information. That is, HSPP may also confer informational advantages regarding things like university culture and practical knowledge about majors, disciplines, housing, and social events. Students from more socioeconomically advantaged backgrounds are typically also at an informational advantage, as more advantaged students have better information about the college application process (Engle et al., 2006), higher college enrollment (Sandefur et al., 2006), and higher retention rates when they do attend (Soria & Stebleton, 2012). Students from working-class backgrounds often do not know anyone when they arrive on campus, which can deprive them of potentially useful information about their best fit for housing, what classes to take, and what social events to attend (Armstrong & Hamilton, 2013). As such, while HSPP does not guarantee belonging or better access to information, it provides students with a bridge to these outcomes and, if realized, may contribute to both long- and short-term college success.

*Peer Prevalence Among First-Generation Students.* If HSPP does independently shape students’ outcomes, we hypothesized that it should be especially important for students from historically marginalized and underrepresented backgrounds, particularly first-generation college students (i.e., students whose parents do not hold a 4-year college degree).<sup>1</sup> First-generation college students tend to have lower college retention rates than those of continuing generation students (Soria & Stebleton, 2012). Evidence indicates that at least part of this can be explained by a mismatch between the dominant norms on college campuses and the norms and experiences that more commonly characterize these students’ upbringing. Norms on college campuses prize independence, autonomy, and individual merit, but

first-generation students, who are more likely to have working class and blue-collar backgrounds, tend adhere to relatively interdependent, communal, and familial norms (Stephens, Fryberg, et al., 2012). This mismatch has a number of downstream implications for these students, including heightened cortisol (a stress hormone) and negative emotions when college norms are framed in ways that emphasize independence (“you are on your own,” Stephens, Townsend, et al., 2012). However, research has found increased performance among first-generation students who actively affirmed their own independent values (Harackiewicz et al., 2014; Tibbetts et al., 2016), and thereby moved their self-construal more in line with the independent context. Similarly, in an influential review, Guiffrida (2006) argued that familial and home connections may be particularly important for underrepresented students, such as first-generation college students. That is, in contrast to Tinto’s (1993) notion that college students benefit from “breaking away” from their family and home cultures, minority students may experience greater benefits from maintaining their prior familial and cultural connections in college.

We reasoned that for first-generation college students, HSPP could help mitigate cultural mismatch. Having a network of peers in place at matriculation may help these students feel a sense of interdependence with similar others while navigating the relatively independent university norms. Having peers from one’s high school may help minority students integrate where they came from (i.e., high school and regional culture and identity) with their college identity. In one study, only first-generation college students benefitted from the social capital afforded by the weak ties of Facebook friends (Wohn et al., 2013). That is, having access to people who could provide informational support about going to college (i.e., bridging social capital) was associated with higher expectations of success in college. First-generation students also show particularly strong benefits from connections to institutional agents in college—that is, nonkin members of higher status positions who can provide social and informational support (Moschetti & Hudley, 2015; Stanton-Salazar, 1997), and they tend to benefit disproportionately from high perceived peer support (Dennis et al., 2005; for a review, see Mishra, 2020).

### *Present Research*

Below we report our efforts to isolate the effect of HSPP as (a) an outcome of students’ academic and demographic background and (b) a predictor of two indicators of college achievement: performance in large introductory STEM courses and college retention. These two academic outcomes were selected deliberately. First, we focused on introductory STEM performance because of our concern with how HSPP would affect students during these early, challenging courses, many of which were very large (over 200 students) and may make students feel relatively anonymous. The bonding and

bridging social capital provided by higher HSPP may be especially useful in these classes, such that students with high HSPP may get higher grades than their lower HSPP peers.

Second, to understand the potential effects of initial social capital on long-term college outcomes, we examined students' retention at the university. Given that students' ability to establish social connections and intellectual fit following the transition to college are critical to individual students' college retention, students who arrive on campus with high HSPP may arrive with a built-in advantage for retention in college. First-generation college students and students from underrepresented backgrounds commonly report feelings of isolation and alienation following the transition to college (Richardson & Skinner, 1992). HSPP may be especially beneficial for such underrepresented students.

In summary, we hypothesized that

**Hypothesis 1:** Having high HSPP would be associated with demographic categories associated with societal privilege and overrepresentation, including being White, continuing generation, and from more affluent high schools.

**Hypothesis 2:** Above and beyond these demographic variables, students who started at the university with high HSPP scores would have better academic performance and higher retention compared with students with low HSPP scores.

**Hypothesis 3:** The effects of HSPP would differ as a function of students' first-generation status or racial/ethnic minority status. Since students in these groups tend to have lower belonging and access to information, a higher number of ties from high school (high HSPP) may be particularly helpful for students from these subgroups.

## Method

### *Sample and Data Sources*

The sample includes data from 43,240 undergraduate students who were enrolled at a large, public research university in the mid-Atlantic portion of the United States between the fall of 2010 and the spring of 2018 and who attended high schools in the United States. Demographic and academic data on this sample of students were provided by the university through a larger project exploring disparities in undergraduate achievement and evaluating interventions specifically in math and science. Thus, to be in the data set, students must have taken at least one biology, chemistry, math, physics, or psychology course. The student-level demographic and administrative data were merged with geographic information obtained from the U.S. Census Bureau (2017) and high school-level data obtained from the U.S. Department of Education's (2017a, 2017b) Common Core

of Data (CCD) and Private School Universe Survey (PSS). Tables 1 and 2 include a description of the sample.

### *Institutional Context*

The university under study is a large, state-related research-intensive university with 19,000 undergraduates (and 10,000 postgraduates). Among undergraduates, approximately 53% of the student body are women, 17% are members of racial/ethnic minority populations, and 9% are first-generation students. The university is moderately selective (60% acceptance rate), has a freshman-to-sophomore retention rate of 93%, a 4-year graduation rate of 70%, and a 6-year graduation rate of 83%. Out of students who do not graduate in 6 years, most transferred to a different institution (73%).

### *Measures*

*High School Peer Prevalence.* This study operationalizes HSPP as the estimated number of students from each individual student's high school concurrently enrolled at the university, assuming a 4-year college completion timeline, when the student matriculates into the university. HSPP is identified using the university's administrative data with students' high school identification numbers and semester of enrollment. The smallest HSPP score in the data set is one, indicating that the student was the only one from their high school who enrolled at the university under study during the same semester and there were no other students from their high school currently enrolled. The largest HSPP score is 329 students. It should be emphasized that since the data set is limited to those who took a math or science course, the peer prevalence variable only captures other students who are also in the data set and is therefore may slightly underestimate the HSPP, since students excluded from the database will not be included in the count. However, general education requirements at the university require students to take at least three natural science courses, and one behavioral science course, thus, the data set captured the vast majority of undergraduates at the university during the time period examined. Furthermore, given the high concentration of students with low HSPP, the analyses use a natural log transformation of the variable to linearize the relationship between students' HSPP and the dependent variables (Cohen et al., 2003).

*Introductory Course Grade.* Our examination of the effect of HSPP on course grades considered grades of introductory courses in biology, chemistry, math, physics, and/or psychology. There are 29,957 students in this analysis, which is about 70% of the students in the data set. The introductory course grade analysis relies on a stacked panel data set, so each student can contribute multiple observations to the analysis if they took more than one of the courses examined

TABLE 1

*Description of the Subset of Students Examined for Analyses on Course Grades (Imputed Estimates)*

Variable	<i>M</i> or %	<i>SD</i>	Min	Max
Level 1: <i>N</i> (observations) = 65,735				
Level 2: <i>N</i> (students) = 29,957				
Number of clusters (high schools) = 3,948				
Course grade	2.65	1.07	0.0	4.0
Gender				
Female	51%			
Male	49%			
Race/ethnicity				
Asian	12%			
Black	6%			
Latinx	3%			
Multiracial	4%			
Other	0.2%			
White	75%			
Parent education				
High school or below	14%			
College or higher	86%			
Distance from home to college (miles)	217.25	346.97	0.11	7,013.62
High school GPA	3.91	0.45	0.00	5.00
SAT verbal/100	6.41	0.78	2.00	8.00
SAT math/100	6.49	0.71	2.00	8.00
Urbanicity				
Large city	8%			
Small city	6%			
Suburb	65%			
Town	7%			
Rural	14%			
Size of high school	1,326	760	0	10,559
High school eligible for FRPL	23%	18%	0%	100%
High school that is non-White	20%	20%	0%	100%
High school peer prevalence	44	63	1	328
Enrolled as a first-time freshman	82%			
Repeated the course	9%			
Course taken in fall semester	66%			
Course taken in spring semester	31%			
Course taken in summer semester	3%			
Introduction to Biology	19%			
Introduction to Chemistry	22%			
Analytic Geometry and Calculus I	21%			
Introduction to Physics	11%			
Introduction to Psychology	27%			

*Note.* FRPL = free or reduced-price lunch.

in the models. A multilevel modeling framework is used to account for this nesting; this is described in more detail in the analytic approach. Course grades are measured on a 4-point scale with an F being equal to 0 and an A or A+ being equal to 4.0.

*Retention.* The measure of retention was a dichotomous indicator indicating whether a student received a bachelor's degree from the university under study or, in the case of students currently enrolled at the university during the time of analysis, remained enrolled as of spring 2018. All other

TABLE 2  
*Description of the Full Sample (Imputed Estimates)*

Variable	<i>M</i> or %	<i>SD</i>	Min	Max
<i>N</i> = 43,240				
<i>N</i> of clusters (high schools) = 4,792				
Retention	75%			
Gender				
Female	51%			
Male	49%			
Race/ethnicity				
Asian	9%			
Black	6%			
Latinx	4%			
Multiracial	4%			
Other	0.2%			
White	78%			
Parent education				
High school or below	16%			
College or higher	84%			
Distance from home to college (miles)	189.19	321.99	0.11	7,013.62
High school GPA	3.79	0.56	0.00	5.00
SAT verbal/100	6.23	0.95	2.00	8.00
SAT math/100	6.25	0.91	2.00	8.00
Urbanicity				
Large city	8%			
Small city	5%			
Suburb	64%			
Town	7%			
Rural	16%			
Size of high school	1,324	819	0	10,559
High school eligible for FRPL	24%	18%	0%	100%
High school that is non-White	19%	20%	0%	100%
High school peer prevalence	52	69	1	328

*Note.* FRPL = free or reduced-price lunch.

students are nonretained students because they dropped out of the university or, more commonly, transferred to another university. Unfortunately, we are not able to differentiate between students who transferred to another university from those who dropped out because we do not have information on why a student left the university.

*First-Generation Status.* First-generation status is measured with a dummy variable indicating whether the highest level of education achieved by a student’s parent is less than a college degree or a college degree or higher (reference group). This information is derived from an administrative variable indicating whether the student is first generation. In cases where the administrative indicator is missing, first-generation status is calculated using parent education information originally obtained by the university from students’ Free Applications for Federal Student Aid (FAFSA).

*Variables in the Model.* *Gender* is represented with a dummy variable indicating whether the student is coded as a male (reference group) or female. *Race/ethnicity* is represented with dummy variables indicating whether the student is Asian, Black, Latinx, multiracial, White (reference group), or other. “Other” captures students who identify as Native American, Pacific Islander, or other. These groups are consolidated into a single race/ethnicity category due to small sample sizes. For ease in testing Hypothesis 1, we use historically privileged social groups, males and Whites, as the reference groups for gender and race/ethnicity dummy codes, respectively. These variables were represented with dummy codes, which, unfortunately, prevented us from running sub-analyses on the gender nonbinary student population or on the various subtypes of multiracial identities on campus.

The university under study is a public university that draws many students from the immediate surrounding areas.

Thus, those who are from areas close to the university are also likely to have larger HSPP, so analyses include the *distance from students' home zip code to the university under study zip code* in order to control for other factors related to proximity to home that may affect students' academic performance and retention. The distance variable is calculated based on home zip codes provided by the university under study and corresponding longitude and latitude coordinates retrieved from U.S. Census Bureau (2017) data. The distance variable is created using NEARSTAT (Jeanty, 2010) in Stata 15.0 and kilometers are converted to miles for analysis and then log transformed in order to linearize the relationship between distance from the students' home to college and the dependent measures (Cohen et al., 2003).

A student's academic performance in high school is likely to be related to a student's academic performance and retention in college, so high school academic performance is controlled for in all analyses. This includes students' *high school grade-point averages* (GPA). The high school GPA data provided by the university under study is not standardized, so GPAs above 5.0 were removed from analyses due to the fact that the majority of students' high school GPAs fall within the 0.0 to 5.0 range. Thus, those with high school GPAs above 5.0 likely went to a high school that did not calculate GPAs on a 4.0 scale, making their high school GPAs less comparable to the majority of students in the sample.

The analyses also include students' Scholastic Aptitude Test (SAT) scores or American College Testing (ACT) scores. For students with ACT scores instead of SAT scores, ACT English and math scores are standardized to the SAT scale using the College Board (2018) ACT/SAT concordance tables to create an equivalent *verbal SAT* score and a *math SAT* score for each student.

Characteristics of a student's high school including enrollment size, racial/ethnic, and socioeconomic composition, as well as urbanicity are likely to influence how many students from the high school matriculate to the university under study, which is directly implicated in HSPP. To capture this source of variability, student-level files from the university under study were merged with a students' high school identification numbers in the U.S. Department of Education's (2017a, 2017b) CCD and PSS. The CCD collects annual data on U.S. public school characteristics and the PSS collects bi-annual data on U.S. private school characteristics. Student-level data were merged with high school data for the academic year prior to the academic year the student matriculated to the university under study, or in the case of students who attended private high schools, 1 or 2 years prior depending on whether the PSS was administered in the previous year. The variables from the CCD and PSS included as covariates in the analyses are the *size of the high school's student population*, the *percentage of the high school that is non-White*, the *percentage of the high school that is eligible for free or reduced-price lunch*, and the *high*

*school's urbanicity*. Urbanicity is measured through a series of dummy variables for large city, small city, suburb (reference group), town, and rural. A correlation table for all variables included in the analyses is included in Supplemental Materials (available at <https://osf.io/4ahs5/>).

The analyses examining introductory course grade controlled for additional course characteristics using dummy variables including whether the student enrolled as a first-time, first-year student, whether the student repeated the course, and what semester the student took the course: spring, summer, or fall (reference group). Each of the following introductory courses is also included in the model as a dummy variable: Introduction to Biology, Introduction to Chemistry, Analytic Geometry and Calculus 1, Introduction to Physics, and Introduction to Psychology (reference group).

#### Missing Data

There were 19,997 students, roughly 46% of the sample, who had complete data for all variables from all data sources. The missing data is largely attributed to missing information in the CCD and PSS, as well as missing high school identification crosswalk codes in the administrative data provided by the university under study that allow the administrative data to be merged with the CCD and PSS. Additionally, almost 7,000 students in the data set have no first-generation identification either from the administrative indicator or the FAFSA, contributing to the amount of missing data. In order to run analyses on the full sample of 43,240 students, missing data were imputed using chained equations to create 20 complete data sets in Stata 15.0 (Royston, 2004, 2005). The study relies on imputation for missing data to obtain the least biased estimates. If the study only analyzed data from the 46% of students with complete data, the results would not be representative since this subsample systematically excludes important groups of students (e.g., students who did not complete the FAFSA). For simplicity, we only present the results with imputed data; however, comparisons with unimputed data yielded largely consistent results.

#### Analytic Approach

All the analyses for hypothesis testing included the same predictor variables based on student-level and high school-level information. We ran a series of regression analyses in Stata 15.0 to examine the associations between students' demographic, academic, and high school backgrounds and HSPP (Hypothesis 1), and how HSPP in turn predicted their introductory STEM course grades and retention in college (Hypothesis 2). Analyses also examined possible interactions between students' HSPP and students' gender, racial/ethnic minoritized status, and first-generation status on the college outcomes in Hypothesis 2 (Hypothesis 3). First-generation status is the only characteristic of the three tested

in Hypothesis 3 that yielded a significant interaction effect and is therefore the only interaction discussed in the results. Overall, Hypothesis 1 highlights which student and high school characteristics predict rates of HSPP in college, while Hypothesis 2 and Hypothesis 3 highlight the role of HSPP in predicting student outcomes in college.

Each model testing Hypothesis 1, Hypothesis 2, and Hypothesis 3 includes a cluster adjustment for students' high school to account for correlated standard errors of students from the same high school. There are a total of 3,948 high schools in the analyses predicting STEM course grades and 4,792 high schools in the analyses predicting HSPP and retention. Furthermore, the analyses examining STEM course grades used a two-level model to account for students who were enrolled in more than one of the STEM courses included in the analyses.

*HSPP.* To examine predictors of HSPP, we conducted a linear regression analysis including all noted background variables as predictors.

*STEM Performance.* As noted, we examined the association between students' HSPP and introductory course grades in biology, chemistry, math, physics, and psychology. We then examined interaction effects to test if the main effects differed as a function of first-generation status. Finally, we tested three-way interactions to examine if the main effects or two-way interactions differed across courses. As previously mentioned, we used a two-level model with a random intercept for student (to account for the nesting of courses within students). A multilevel approach was chosen to account for the fact that the same student may have taken more than one introductory course; for example, certain majors require both Introduction to Biology and Introduction to Chemistry, so students who took both have multiple observations in the analyses. The imputed two-level models included 65,735 observations and 29,957 students, which is an average of about 2.2 observations per student.

*Retention.* Analyses on retention examined the relationship between students' HSPP and retention at the university under study using logistic regression due to the dichotomous outcome variable of retention. We then examined interaction effects to test if the main effects of HSPP differed as a function of first-generation status. Although here we report analyses focusing broadly on retention, similar results were found for first-generation students in a narrower examination with college graduation as the outcome (see online Supplemental Materials).

## Results and Discussion

### *Predictors of HSPP*

Analyses showed that HSPP was multiply and independently determined by a host of demographic and high school

variables. These variables paint a picture of HSPP as being partly rooted in societal privilege. Students with high HSPP were more likely than students with low HSPP to come from a relatively close, suburban, largely White, and affluent high school. They also tend to have a high GPA in high school, and are more likely to be White, male, and have at least one parent with a college degree. As such, evidence largely supported the first hypothesis that HSPP would be associated with historically overrepresented, privileged background characteristics (see online Supplemental Materials).

To examine predictors of HSPP (Hypothesis 1), we treated HSPP as the outcome variable and all the demographic variables as predictors. Here we note the significant effects by reporting their unstandardized coefficients. Please see Table 3 for details on the model, including coefficients, and standard errors. Namely, high HSPP was predicted by being a man ( $B = .066$ , standard error [ $SE$ ] =  $.019$ ,  $p < .001$ ), continuing-generation ( $B = .08$ ,  $SE = .022$ ,  $p < .001$ ), and, using Whites as the reference group, not Black ( $B = -.34$ ,  $SE = .050$ ,  $p < .001$ ), not Latinx ( $B = -.225$ ,  $SE = .040$ ,  $p < .001$ ), and not multiracial ( $B = -.113$ ,  $SE = .035$ ,  $p < .01$ ). HSPP was higher among students that had higher high school GPA ( $B = .298$ ,  $SE = .039$ ,  $p < .001$ ), and, with suburban schools as the reference group, were not from schools in a small city ( $B = -.591$ ,  $SE = .098$ ,  $p < .001$ ), town ( $B = -.632$ ,  $SE = .151$ ,  $p < .001$ ), or rural setting ( $B = -.400$ ,  $SE = .101$ ,  $p < .001$ ). Students with higher HSPP also tended to come from high schools whose student body had a low percentage of both non-White students ( $B = -1.222$ ,  $SE = .205$ ,  $p < .001$ ) and free/reduced lunch recipients ( $B = -1.046$ ,  $SE = .214$ ,  $p < .001$ ).

There was also an unexpected association involving students' SAT/ACT verbal test scores, as the results showed that having higher scores was associated with slightly lower HSPP ( $B = -.109$ ,  $SE = .016$ ,  $p < .001$ ). In addition, as expected, the log of distance from students' home zip code to campus was a significant predictor, with HSPP declining as the log-distance from campus increased ( $B = -.59$ ,  $SE = .029$ ,  $p < .001$ ). Also as expected, students who come from high schools with larger enrollment had higher HSPP ( $B = .0005$ ,  $SE = .0001$ ,  $p < .001$ ).

### *STEM Performance in Gateway Courses*

With an understanding of the sources of HSPP, we next sought to examine whether it predicted college outcomes above and beyond these background factors. To address Hypothesis 2 and Hypothesis 3, we first examined how HSPP influenced students' grades in gateway natural science introductory courses: Introductory Psychology, Introductory Math (Analytic Geometry and Calculus I), Introductory Chemistry, Introductory Biology, and Introductory Physics for Engineers. The results yielded several main effects and interactions (see Table 4). Consistent with Hypothesis 2, there was a significant overall main effect of HSPP on course



Table 3.  
Variables Predicting HSPP (with cluster adjustment for high school)

	(Model 1) HSPP	(Model 2) HSPP	(Model 3) HSPP
Women (Reference=Men)	-0.030* (0.015)	-0.028 (0.018)	-0.067*** (0.019)
First-Generation	-0.245*** (0.040)	-0.201*** (0.032)	-0.076*** (0.022)
Black (Reference=White)	-0.870*** (0.061)	-0.790*** (0.065)	-0.342*** (0.051)
Latinx (Reference=White)	-0.398*** (0.046)	-0.349*** (0.046)	-0.225*** (0.040)
Asian (Reference=White)	0.064 (0.059)	0.0037 (0.052)	0.022 (0.039)
Multiracial (Reference=White)	-0.263*** (0.043)	-0.251*** (0.043)	-0.114** (0.035)
Other Race (Reference=White)	-0.051 (0.144)	-0.050 (0.142)	0.021 (0.122)
Distance in miles from home zip code to main campus (log transformed)	-0.583*** (0.034)	-0.590*** (0.034)	-0.589*** (0.029)
High School GPA		0.207*** (0.045)	0.298*** (0.039)
SAT Verbal		-0.089*** (0.019)	-0.109*** (0.016)
SAT Math		0.106*** (0.024)	0.010 (0.019)
Large City (Reference=Suburb)			-0.154 (0.104)
Small City (Reference=Suburb)			-0.590*** (0.098)
Town (Reference=Suburb)			-0.632*** (0.151)
Rural (Reference=Suburb)			-0.401*** (0.101)
High School Enrollment			0.0004*** (0.0001)
Percent of High School Eligible for Free or Reduced Price Lunch			-1.041*** (0.213)
Percent of High School that is Non-White			-1.224*** (0.206)
_cons	5.565*** (0.178)	4.697*** (0.244)	5.114*** (0.204)
N	43240	43240	43240
# of Clusters (high schools)	4,792	4,792	4,792

Note. Standard errors are in parentheses. + $p < .10$ , \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

TABLE 4  
Results of Two- and Three-Way HSPP, First-Generation (FG), and Course Interactions on Students' Grades in Introductory STEM Courses

Main effects and interaction	Introductory Biology		Introductory Chemistry		Analytic Geometry and Calculus 1		Introduction to Physics	
	B	SE	B	SE	B	SE	B	SE
HSPP	.012 <sup>†</sup>	.006	.024***	.006	.023**	.007	.038***	.010
FG status	-.050	.041	-.021	.038	-.072	.047	-.139*	.063
HSPP × FG status	-.000	.003	-.000	.003	.003	.003	.008*	.004

Note. The reference category for the courses was Introduction to Psychology. See online Supplemental Materials for full regression table. HSPP = high school peer prevalence; STEM = Science, Technology, Engineering, and Mathematics.

<sup>†</sup> $p < .10$ . \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

grades before accounting for interactions ( $B = .015$ ,  $SE = .005$ ,  $p = .002$ ). Higher HSPP independently predicted higher average course grades across first-year STEM courses. Further analyses examined how the slope of HSPP varied across courses and as a function of first-generation status (i.e., by testing  $HSPP \times$  course and  $HSPP \times$  first-generation interactions). These analyses revealed the overall main effect was qualified in several ways.

First, we found the main effect was stronger in some courses than in others. In particular, the effect of peer prevalence was not a significant predictor of Psychology course grades ( $B = -.002$ ,  $SE = .006$ ,  $p = .741$ ). However, it was a significant predictor of grades in each of the other courses. HSPP predicted higher Biology grades ( $B = .012$ ,  $SE = .006$ ,  $p < .05$ ), Chemistry grades ( $B = .024$ ,  $SE = .006$ ,  $p < .001$ ), and math grades ( $B = .023$ ,  $SE = .007$ ,  $p < .01$ ; with

TABLE 5  
Results of HSPP and First-Generation (FG) Interaction on Student Retention

Main effect and interaction	B	SE	p	OR	95% CI OR
HSPP	-.035	.03	.207	.97	[0.914, 1.020]
FG status	-.510	.08	<.0001	.60	[0.513, 0.703]
HSPP × FG status	.114	.02	<.0001	1.12	[1.069, 1.176]

Note. HSPP = high school peer prevalence.

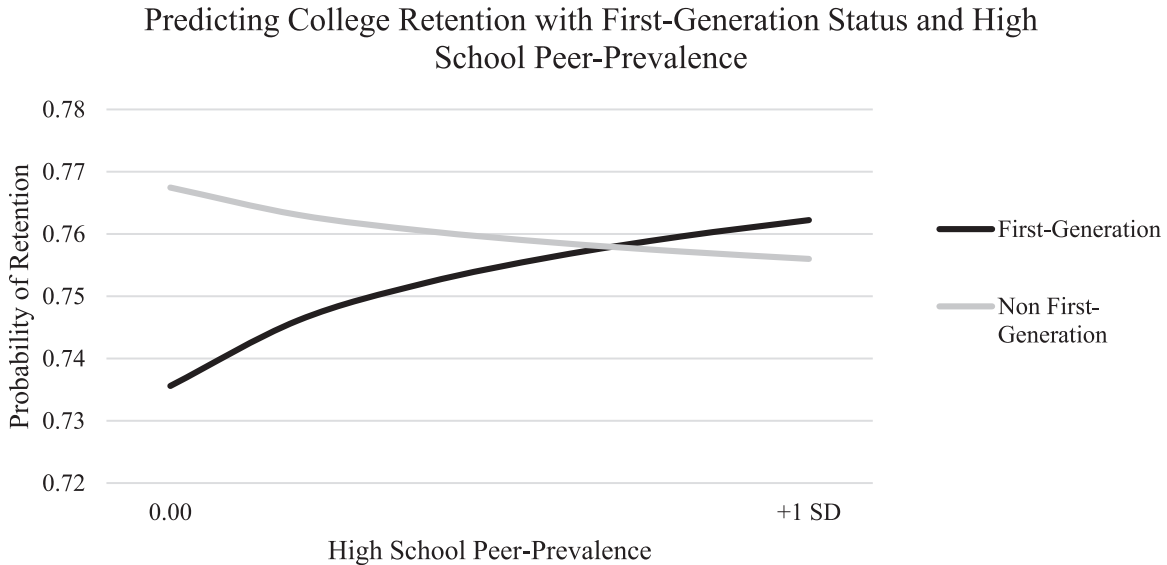


FIGURE 1. High school peer prevalence (HSPP; 0 to +1 SD of the mean) and first-generation status interaction on probability of being retained at the university, defined as either having graduated or currently enrolled in the current semester in which the data were collected.

Psychology grades as the reference course for each effect). There was a .011-point increase in Biology course grade, a .024-point increase in Chemistry course grade, and a .023-point increase in Math course grade for every additional log-unit change in peer prevalence. There was also a significant effect of HSPP on Physics grades ( $B = .043$ ,  $SE = .010$ ,  $p < .001$ ), but this effect was further qualified by first-generation status (consistent with Hypothesis 3).

Namely, the final step of the analysis was to test for three-way interactions, which uncovered an  $HSPP \times$  First-generation  $\times$  Physics interaction on course grade ( $B = .008$ ,  $SE = .004$ ,  $p < .05$ ). For continuing-generation students, one log-unit change in HSPP was related to a .04-point increase in their Physics grade, compared with their Psychology grade ( $B = .038$ ,  $SE = .010$ ,  $p < .001$ ). First-generation college students, however, benefited more from having higher HSPP in Physics as compared with Psychology. For these students, having a higher HSPP was related to a .08-point increase in their Physics grades ( $B = .078$ ,  $SE = .025$ ,  $p < .01$ ).

These findings indicated that while peer prevalence was a significant predictor for all students overall, it was

particularly strong predictor in Biology, Chemistry, and Math and for first-generation college students in Physics. This latter finding was consistent with Hypothesis 3, although it is notable that we only observed support for it in one course. We note that these four courses were all relatively difficult courses, each with an average course grade of C+. Psychology, by contrast, had an average course grade of B. See Appendix Table A1 for full regression model output.

#### Retention

We next examined the effect of HSPP on students' likelihood of staying and graduating at the university (Table 5). To analyze student retention, we conducted a binary logistic regression (0 = not retained; 1 = currently still enrolled as of spring 2018 or graduated). The analyses did not support Hypothesis 2, as there was no overall main effect of HSPP on retention. However, in line with Hypothesis 3, analysis revealed a  $HSPP \times$  First-generation status interaction on retention ( $B = .114$ ,  $SE = .024$ ,  $p < .001$ , odds ratio [OR] = 1.12, 95% CI [.219, .554]; Figure 1). In continuing generation

students, peer prevalence did not have a significant effect on retention ( $B = -.035, SE = .028, p > .10$ ). However, among first-generation students, peer prevalence positively influenced retention ( $B = .079, SE = .024, p = .001$ ). This interaction is plotted in Figure 1. Among first-generation students, a one log-unit increase was related to an almost 1% increase in likelihood of retention; for continuing-generation students, however, a one-unit increase was related to a slight decrease (<1%) in retention. See Appendix Table A2 for full regression model output.

To summarize across both outcomes, both Hypothesis 2 and Hypothesis 3 found support, albeit unevenly across outcomes. Analyses on course performance found support for a main effect of HSPP on STEM course grades, in support of Hypothesis 2. However, there was only mixed support for Hypothesis 3 on introductory course grades, as HSPP was particularly important for first-generation students in Physics courses—one of the more difficult courses under study—but there was no such interaction effect involving first-generation status on grades in other courses. Analyses on college retention did not find support for Hypothesis 2 but did find support for Hypothesis 3. Thus, there was no overall effect of HSPP on retention, but HSPP did predict increased retention among first-generation students. Across all analyses, including additional exploratory analyses not reported here, we found very little support for benefits of HSPP as a function of gender or race/ethnicity. As such, for Hypothesis 3, only effects on first-generation students were observed.

### General Discussion

Results showed that a simple metric, derived from the prevalence of one's high school peers on campus, was predicted by variables associated with societal privilege and, in turn, significantly predicted students' introductory STEM course grades. First, in support of Hypothesis 1, HSPP was higher among students who were White, male, continuing-generation, suburban, and had a high GPA from high schools that were relatively White and affluent (i.e., low percentage of free/reduced lunch). Second, in support of Hypothesis 2, above and beyond these demographic and academic variables, peer prevalence predicted higher course grades in difficult introductory STEM courses, including biology, chemistry, and math, but not in an easier introductory course (as judged by average course grade), Psychology. Finally, in support of Hypothesis 3, first-generation college students' HSPP also predicted higher grades in Introductory Physics and higher rates of college retention. In practical terms, this latter finding indicated that the average first-generation student needed 8 to 9 high school peers on campus to reach the same rate of retention (76.7%) as the average continuing generation student (see online Supplemental Materials for additional output tables).

Together, the results support the idea that a relatively minimalist, theory-driven measure of HSPP can reveal one

mechanism by which societal privilege supports college achievement. Although the predictive effects of HSPP were relatively small, we argue the effects are practically significant, as HSPP predicted socially important and difficult-to-influence college outcomes (see Prentice & Miller, 1992). However, while analyses found benefits of HSPP among first-generation college students on Physics grades and college retention, we saw very little evidence of benefits of HSPP as a function of students' race/ethnicity or gender. In the space below, we further unpack the findings, discuss possible explanations for the overall pattern of results, and explore implications for practice.

### *Benefits of HSPP*

Navigating the transition to 4-year universities is challenging for many students. Students from historically marginalized social groups, such as first-generation college students, are often at a disadvantage, as they are more likely contend with doubts about belonging (Walton & Cohen, 2007) and have lower social access to useful college-going information (Armstrong & Hamilton, 2013). First-generation students also tend to experience cultural mismatch, with the relatively interdependent norms of their upbringing being incongruent with the relatively independent norms of university culture (Stephens, Fryberg, et al., 2012). Students from more affluent and educated backgrounds, by contrast, are normative for the university and are less likely to experience a lack of belonging stemming from their demographic or high school background (Layous et al., 2017). We posit that having peers from one's high school present at the university provides students with social capital to help navigate the transition and thrive in college. These benefits may be psychological (increased sense of belonging and connection) and informational (increased access to valuable academic and social knowledge).

As noted, results revealed a significant main effect of peer prevalence on introductory STEM course grades, an effect that was especially strong in difficult STEM gateway courses (as judged by average GPA) compared with Introductory Psychology. This is in line with evidence indicating that large, introductory STEM courses pose specific challenges to students (Matz et al., 2018), including a sense of anonymity and novelty around the size and structure of the learning environment (Scott et al., 2017), which high HSPP may help students manage. In one of the more difficult courses, Introductory Physics, first-generation students showed a particularly strong benefit of HSPP on grades. Although speculative, perhaps the relative difficulty of this course made having a social network available for social and informational support particularly beneficial.

The finding that HSPP also predicted retention for first-generation college students suggests that the benefits of HSPP may persist after the first year. Given evidence that first-generation college students face hurdles to a successful

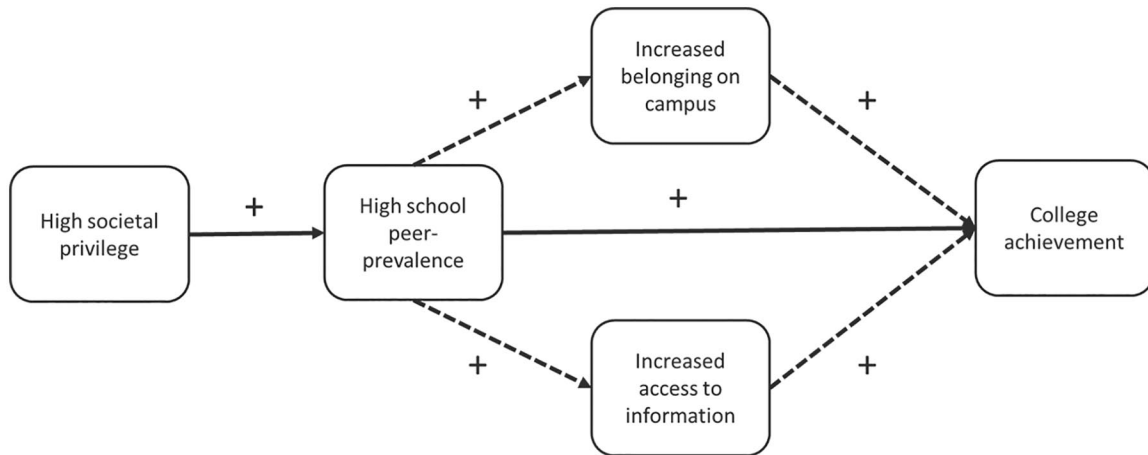


FIGURE 2. Schematic of proposed sequence for how high school peer prevalence (HSPP) partially mediates the effect of societal privilege on college achievement.

Note. Solid paths indicate links that were supported in the present research. Dotted paths indicate theorized links that were not assessed in the present research.

transition because of cultural mismatch (Stephens, Townsend, et al., 2012), perhaps arriving at college with a social network in place aided the transition to college, placing first-generation students on an improved long-term trajectory resulting in higher retention. As first-generation students encountered the independent norms that pervade university culture, those with higher HSPP may have been buffered from the cultural mismatch by virtue of their built-in social network. For example, first-generation students with higher HSPP may have been less likely to infer that people from their background do not belong at the university. These are questions to explore in future research, for example, by conducting qualitative interviews with first-generation students to investigate if and how those with high HSPP utilized their former high school network in college. Notably, while all students seemed to benefit from HSPP in difficult, introductory STEM courses shortly after transitioning to college, only first-generation students showed long-term benefits of HSPP on retention. As such, another question for future research is whether first-generation and continuing-generation students use their social networks in similar or different ways over time.

#### *Effects Across Demographic Groups*

Benefits of HSPP were found for first-generation students but not for racial/ethnic minority students or women. One factor that makes first-generation status different from these other categories is its relative invisibility. That is, outward appearances generally mark members of gender and racial/ethnic minority groups, whereas first-generation status does not have widely known physical markers. Research on social stigma has long recognized the importance of the visibility versus invisibility distinction (Goffman, 1963; Henning et al., 2019). Concealable

stigmatized identities (CSI; Crocker et al., 1998) refer to negatively valued social identities that are not associated with physical characteristics or outward markers (such as having a mental illness, being a victim of abuse, or having health complications).

Research indicates that individuals contending with a CSI commonly face less overt prejudice and fewer negative social experiences owing to their identity than individuals who cannot conceal their stigma (Pachankis, 2007). However, the lack of visibility also has social costs. For one, not being able to identify people with the identity (e.g., first-generation students) means that stigmatized group members may have a harder time finding and connecting with one another. In this way, first-generation students are deprived of readily observable indicators that others in their environment are “like them” and possibly experiencing a similar reality (unlike for racial or gender groups; see Binning & Unzueta, 2013; Greenaway & Turetsky, 2020; Murphy et al., 2007). HSPP may provide such students with an alternative means of knowing there are others who are like them on campus. Moreover, if HSPP predicts benefits for one CSI, first-generation status, it may also be beneficial for students contending with other CSIs (e.g., learning disabilities). This is another question for future research.

#### *Limitations*

While the measure of HSPP was readily available and easy to calculate, we have not yet examined the theorized mediators of HSPP on college achievement. Figure 2 depicts a schematic of theorized links among the variables of interest. Several links were inferred from the literature and were not directly addressed in the present research.

Most notably, HSPP is not a direct measure of either belonging or access to information. As such, students' belonging and access to information are proposed as mediators of HSPP on student outcomes, but additional research is necessary to examine if either or both do, in fact, explain HSPP's effects. Similarly, by relying on administrative data, we did not actually measure students' perceptions of their peers (e.g., whether they view high school peers positively or negatively), which would help shed light on how HSPP has an influence, such as whether it was just as beneficial to have many weak social ties from high school versus fewer strong social ties.

Although we strived to control for major potential confounds with HSPP, the correlational nature of the data opens the possibility that other unmeasured variables contributed to the results. One strength of the findings, however, is that HSPP did not simply show a main effect on course performance; it also operated on Physics grades and retention in a way that was consistent with our theorizing and previous research on the experiences of first-generation college students. An imperative next step is to examine whether this interaction effect between HSPP and first-generation status replicates in other social university contexts, for example, in contexts with higher and lower percentages of first-generation college students.

Another limitation is that the present work did not consider whether and to what extent college choice may have been driven by HSPP. For example, when choosing where to apply or accept, did students prefer colleges where their high school peers were attending (see Hossler & Gallagher, 1987)? Is their tendency to enroll in the summer after acceptance (see Castleman & Page, 2020) driven in part by HSPP? The presence of high school peers could be especially attractive to more extroverted or socially oriented students (see Armstrong & Hamilton, 2013). By contrast, students who choose to attend schools where they have solo status (e.g., because they choose a more prestigious, distant school) might be more prone to loneliness or belonging uncertainty. Having a more complete picture of the role of HSPP on college choice would help inform the implications of the present findings for practice.

Finally, we note that the present findings are intended to generalize to other large, 4-year universities. Different patterns may unfold in other types of higher education institutions, such as community colleges, where students may have particularly high levels of high school connections both at their institution and in their home community.

#### *Implications for Practice*

The results indicate one process by which societal privilege can create social ecologies in which privilege

is reproduced. Having high school friends independently predicted course grades in first-year STEM gateway courses. However, the results also show that HSPP may be used as a tool to promote equity, as supported by improved retention among first-generation students with high HSPP. With the caveats that the present findings may differ across university contexts and that researchers should investigate such differences before initiating policy changes or interventions (see Binning & Browman, 2020; Harackiewicz & Priniski, 2018), the present insights could inform both admissions and student-support processes. For example, the creation of HSPP is likely to be at least partly driven by admissions processes that favor students from particular high schools (e.g., high-SES high schools) over others. As such, the present findings argue for a critical examination of how admissions processes produce HSPP and how it might be mitigated or leveraged to improve outcomes for all students. It is not simply important for students to build social connections on campus; they also benefit from the connections they already have when they arrive.

On the student support side, our results suggest that while all students may benefit early in college from HSPP, it may be especially important for the retention of first-generation college students. Students who arrive on campus with no known social connections from their high school may be candidates for additional social support and resources, particularly if they are first-generation college students. Students' lack of preexisting social networks when they arrive may be strong candidates for additional outreach and support. For example, by connecting students who share solo status, administrators can help create a network of students who share a consequential identity (see also Petty, 2014).

#### **Conclusion**

This research defined and operationalized a novel, theory-driven measure for understanding college success in STEM gateway courses and college retention among first-generation college students. It did so by using a large, comprehensive data set to examine both the predictors of HSPP and how the effects of HSPP may differ across different demographic groups. These results suggest HSPP is one conduit through which societal privilege may produce social circumstances that facilitate positive life outcomes. By recognizing this role of HSPP, efforts may be undertaken to produce those social circumstances among less privileged students to foster equity in higher education.

## Appendix

Table A1.

**Introductory course grade analysis. Mixed effects model with random intercept for student and cluster adjustment for high school, includes students from fall 2010 through spring 2018 who took introductory courses in Biology, Chemistry, Math (Calculus), Physics, and/or Psychology at the main university campus**

	+ High School Performance/ Characteristics	+ Course Characteristics	+ Cohort and wave	+ firstgen interaction	+ 2 way interactions	+ 2 way and 3 way interactions
Women	0.117*** (0.011)	0.044*** (0.010)	0.045*** (0.010)	0.045*** (0.010)	0.045*** (0.010)	0.045*** (0.010)
First-Generation	-0.219*** (0.017)	-0.048*** (0.014)	-0.047*** (0.014)	-0.059* (0.030)	-0.035 (0.033)	-0.015 (0.041)
Black	-0.460*** (0.028)	-0.083*** (0.023)	-0.077*** (0.023)	-0.077*** (0.023)	-0.078*** (0.023)	-0.078*** (0.023)
Latinx	-0.257*** (0.033)	-0.076*** (0.025)	-0.072** (0.025)	-0.072** (0.025)	-0.072** (0.025)	-0.072** (0.025)
Asian	0.092*** (0.018)	0.010 (0.015)	0.009 (0.015)	0.009 (0.015)	0.009 (0.015)	0.009 (0.015)
Multiracial	-0.088** (0.030)	-0.010 (0.023)	-0.009 (0.023)	-0.009 (0.023)	-0.009 (0.023)	-0.009 (0.023)
Other race	-0.0643 (0.118)	-0.082 (0.095)	-0.082 (0.095)	-0.082 (0.095)	-0.082 (0.095)	-0.082 (0.095)
Log distance to campus	0.009 (0.006)	-0.003 (0.006)	0.006 (0.007)	0.006 (0.007)	0.006 (0.007)	0.006 (0.007)
High school GPA	0.581*** (0.022)	0.568*** (0.021)	0.563*** (0.021)	0.563*** (0.021)	0.563*** (0.021)	0.563*** (0.021)
SAT verbal	0.124*** (0.009)	0.111*** (0.008)	0.113*** (0.008)	0.113*** (0.008)	0.113*** (0.008)	0.113*** (0.008)
SAT quantitative	0.233*** (0.010)	0.276*** (0.009)	0.276*** (0.009)	0.276*** (0.009)	0.276*** (0.009)	0.276*** (0.009)
Large city	-0.045 (0.027)	-0.045+ (0.024)	-0.043+ (0.024)	-0.043+ (0.024)	-0.043+ (0.024)	-0.043+ (0.024)
Small city	-0.044+ (0.025)	-0.044+ (0.022)	-0.038+ (0.023)	-0.038+ (0.023)	-0.038+ (0.023)	-0.0380+ (0.023)
Town	0.008 (0.0242)	0.017 (0.021)	0.025 (0.021)	0.025 (0.021)	0.025 (0.021)	0.025 (0.021)
Rural	-0.0302+ (0.018)	-0.012 (0.016)	-0.006 (0.016)	-0.006 (0.016)	-0.006 (0.016)	-0.006 (0.016)
HS enrollment size	-3.43E-07 (0.00001)	-2.55E-06 (0.00001)	-9.86E-06 (0.00001)	-9.89E-06 (0.00001)	-9.63E-06 (0.00001)	-9.64E-06 (0.00001)
Percent free/red. lunch HS	-0.074 (0.047)	-0.047 (0.041)	-0.031 (0.041)	-0.031 (0.041)	-0.029 (0.041)	-0.029 (0.041)
Percent non-white HS	-0.158** (0.048)	-0.132** (0.041)	-0.118** (0.041)	-0.118** (0.041)	-0.119** (0.041)	-0.120** (0.041)
First-year student		-0.059** (0.022)	-0.055* (0.022)	-0.055* (0.022)	-0.050* (0.021)	-0.050* (0.021)
Repeat enrollment		-1.294*** (0.012)	-1.294*** (0.012)	-1.294*** (0.012)	-1.292*** (0.012)	-1.292*** (0.012)
Spring term		0.079*** (0.009)	0.079*** (0.009)	0.079*** (0.009)	0.077*** (0.009)	0.077*** (0.009)
Summer term		0.417*** (0.022)	0.418*** (0.022)	0.418*** (0.022)	0.418*** (0.022)	0.418*** (0.022)
Intro Biology		-0.625*** (0.009)	-0.625*** (0.009)	-0.625*** (0.009)	-0.651*** (0.018)	-0.651*** (0.019)
Intro Chemistry		-0.509*** (0.009)	-0.508*** (0.009)	-0.508*** (0.009)	-0.572*** (0.017)	-0.572*** (0.017)
Intro Math/Calculus		-0.611*** (0.011)	-0.612*** (0.011)	-0.612*** (0.011)	-0.674*** (0.022)	-0.669*** (0.023)
Intro Physics		-0.737*** (0.016)	-0.737*** (0.016)	-0.737*** (0.016)	-0.855*** (0.026)	-0.841*** (0.027)
HSPPP		0.015** (0.005)	0.015** (0.005)	0.015** (0.006)	-0.002 (0.006)	-0.001 (0.006)
HSPPP_x_firstgen				0.004 (0.010)	0.006 (0.010)	-0.002 (0.012)
HSPPP_x_Bio					0.012* (0.006)	0.012+ (0.006)
HSPPP_x_Chem					0.024*** (0.005)	0.024*** (0.006)
HSPPP_x_Math					0.024*** (0.007)	0.023** (0.007)
HSPPP_x_Physics					0.043*** (0.010)	0.038*** (0.010)
Firstgen_x_Bio					-0.052* (0.026)	-0.050 (0.041)
Firstgen_x_Chem					-0.022 (0.024)	-0.021 (0.038)
Firstgen_x_Math					-0.045 (0.030)	-0.072 (0.047)
Firstgen_x_Physics					-0.048 (0.035)	-0.139* (0.063)
HSPPP_x_firstgen_x_Bio					-0.0003 (0.003)	-0.0003 (0.003)
HSPPP_x_firstgen_x_Chem						0.003 (0.003)
HSPPP_x_firstgen_x_Math						0.008* (0.004)
HSPPP_x_firstgen_x_Physics						-1.485*** (0.074)
_cons	2.675*** (0.032)	-1.440*** (0.065)	-1.517*** (0.073)	-1.516*** (0.073)	-1.482*** (0.074)	-1.485*** (0.074)
N	65735	65735	65735	65735	65735	65735
Total # of students	29,957	29,957	29,957	29,957	29,957	29,957
# of clusters	3,948	3,948	3,948	3,948	3,948	3,948

*Note.* Mixed effects model with random intercept for student and cluster adjustment for high school, includes students from fall 2010 through spring 2018 who took introductory courses in Biology, Chemistry, Math (Calculus), Physics, and/or Psychology at the main university campus. HS = high school; HSPPP = high school peer prevalence; firstgen = first generation; free/red. lunch = free/reduced-price lunch. \*p < .10. \*\*p < .05. \*\*\*p < .01. \*\*\*\*p < .001.

Table A2.

**Retention analysis.** Logistic regression with cluster adjustment for high schools, includes students from fall 2010 through spring 2018 who took introductory courses in Biology, Chemistry, Math (Calculus), Physics, and/or Psychology at the main university campus. Includes students who enrolled but did not receive grades in these courses.

	Demographics	+High School Performance/ Characteristics	+ Cohort and wave	+interaction
Women	0.171*** (0.028)	0.094** (0.030)	0.093** (0.030)	0.091** (0.030)
First-Generation	-0.344*** (0.044)	-0.165*** (0.039)	-0.166*** (0.038)	-0.510*** (0.080)
Black	-0.412*** (0.061)	-0.277*** (0.062)	-0.282*** (0.064)	-0.276*** (0.064)
Latinx	-0.243*** (0.064)	-0.183** (0.064)	-0.186** (0.065)	-0.181** (0.065)
Asian	0.262*** (0.059)	0.137* (0.057)	0.137* (0.057)	0.139* (0.057)
Multiracial	-0.050 (0.063)	-0.036 (0.063)	-0.038 (0.063)	-0.037 (0.063)
Other race	-0.078 (0.277)	-0.070 (0.287)	-0.070 (0.286)	-0.066 (0.286)
Log distance to campus	0.214*** (0.0185)	0.185*** (0.022)	0.177*** (0.020)	0.176*** (0.019)
High school GPA		0.412*** (0.046)	0.416*** (0.043)	0.418*** (0.043)
SAT verbal		0.043* (0.020)	0.042* (0.019)	0.041* (0.019)
SAT quantitative		-0.019 (0.027)	-0.018 (0.027)	-0.019 (0.027)
Large city		-0.066 (0.088)	-0.068 (0.089)	-0.068 (0.088)
Small city		-0.164* (0.065)	-0.173** (0.065)	-0.180** (0.065)
Town		-0.192** (0.070)	-0.201** (0.069)	-0.203** (0.069)
Rural		-0.177*** (0.052)	-0.183*** (0.053)	-0.181*** (0.053)
HS enrollment size		5.4E-06 (0.00002)	1.11E-05 (0.00002)	1E-05 (0.00002)
Percent non-White HS		-0.804*** (0.161)	-0.819*** (0.157)	-0.805*** (0.157)
Percent free/reduced lunch HS		0.453** (0.151)	0.435** (0.147)	0.437** (0.146)
HSPP			-0.014 (0.026)	-0.035 (0.028)
HSPP_x_firstgen				0.114*** (0.024)
_cons	0.213* (0.104)	-1.191*** (0.185)	-1.122*** (0.231)	-1.055*** (0.237)
N	43240	43240	43240	43240
# of clusters	4792	4792	4792	4792

Note. Logistic regression with cluster adjustment for high schools, includes students from fall 2010 through spring 2018 who took introductory courses in Biology, Chemistry, Math (Calculus), Physics, and/or Psychology at the main university campus. Includes students who enrolled but did not receive grades in these courses. HS = high school; HSPP = high school peer prevalence; firstgen = first generation; free/red.pricelunch = free/reduced-price lunch. \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

### Acknowledgments

This work was supported by National Science Foundation Grant #1524575.

### ORCID iD

Kevin R. Binning  <https://orcid.org/0000-0002-5396-4183>

### Note

1. We note that this is imperfect definition that glosses over potentially important distinctions within this demographic category (Nguyen & Nguyen, 2018).

### References

Armstrong, E. A., & Hamilton, L. T. (2013). *Paying for the party*. Harvard University Press. <https://doi.org/10.4159/harvard.9780674073517>

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human

motivation. *Psychological Bulletin*, 117(3), 497–529. <https://doi.org/10.1037/0033-2909.117.3.497>

Binning, K. R., & Browman, A. S. (2020). Theoretical, ethical, and policy implications for conducting social-psychological interventions to close educational achievement gaps. *Social Issues and Policy Review*, 14(1), 182–216. <https://doi.org/10.1111/sipr.12066>

Binning, K. R., Kaufmann, N., McGreevy, E. M., Fotuhi, O., Chen, S., Kalender, Z. Y., Marshman, E., Limeri, L. B., Betancur, L., & Singh, C. (2020). Changing social contexts to foster equity in college science courses: An ecological-belonging intervention. *Psychological Science*, 31(9), 1059–1070. <https://doi.org/10.1177/0956797620929984>

Binning, K. R., & Unzueta, M. M. (2013). Perceiving ethnic diversity on campus: Group differences in attention to hierarchical representation. *Social Psychological and Personality Science*, 4(4), 500–507. <https://doi.org/10.1177/1948550612459771>

Brady, S. T., Cohen, G. L., Jarvis, S. N., & Walton, G. M. (2020). A brief social-belonging intervention in college improves adult outcomes for black Americans. *Science Advances*, 6(18), eaay3689. <https://doi.org/10.1126/sciadv.aay3689>

- Castleman, B. L., & Page, L. C. (2020). *Summer melt: Supporting low-income students through the transition to college*. Harvard Education Press.
- Chen, S., Binning, K. R., Manke, K. J., Brady, S. T., McGreevy, E. M., Betancur, L., Limeri, L. B., & Kaufmann, N. (2021). Am I a science person? A strong science identity bolsters minority students' sense of belonging and performance in college. *Personality and Social Psychology Bulletin*, *47*, 593–606.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates.
- College Board. (2018). *Guide to the ACT®/SAT® Concordance*. <https://www.act.org/content/act/en/products-and-services/the-act/scores/act-sat-concordance.html>
- Crocker, J., Major, B., & Steele, C. M. (1998). Social stigma. In D. T. Gilbert & S. T. Fiske (Eds.), *The handbook of social psychology* (Vol. 2, 4th ed., pp. 504–533). McGraw-Hill.
- Dennis, J. M., Phinney, J. S., & Chuateco, L. I. (2005). The role of motivation, parental support, and peer support in the academic success of ethnic minority first-generation college students. *Journal of College Student Development*, *46*(3), 223–236. <https://doi.org/10.1353/csd.2005.0023>
- Engle, J., Bermeo, A., & O'Brien, C. (2006). *Straight from the source: What works for first-generation college students*. The Pell Institute for the Study of Opportunities in Higher Education. <https://files.eric.ed.gov/fulltext/ED501693.pdf>
- Gee, L. K., Jones, J. J., Fariss, C. J., Burke, M., & Fowler, J. H. (2017). The paradox of weak ties in 55 countries. *Journal of Economic Behavior & Organization*, *133*(January), 362–372. <https://doi.org/10.1016/j.jebo.2016.12.004>
- Goffman, E. (1963). *Stigma: Notes on the management of spoiled identity*. Simon & Schuster.
- Gopalan, M., & Brady, S. T. (2020). College students' sense of belonging: A national perspective. *Educational Researcher*, *49*(2), 134–137. <https://doi.org/10.3102/0013189X19897622>
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380. <https://doi.org/10.1086/225469>
- Greenaway, V. P., & Turetsky, K. M. (2020). Socioecological diversity and inclusion: A framework for advancing diversity science. *Current Opinion in Psychology*, *32*(April), 171–176. <https://doi.org/10.1016/j.copsyc.2019.09.008>
- Guiffreda, D. A. (2006). Toward a cultural advancement of Tinto's theory. *Review of Higher Education*, *29*(4), 451–472. <https://doi.org/10.1353/rhe.2006.0031>
- Hale, C. J., Hannum, J. W., & Espelage, D. L. (2005). Social support and physical health: The importance of belonging. *Journal of American College Health*, *53*(6), 276–284. <https://doi.org/10.3200/JACH.53.6.276-284>
- Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Giffen, C. J., Blair, S. S., Rouse, D. I., & Hyde, J. S. (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. *Journal of Educational Psychology*, *106*(2), 375–389. <https://doi.org/10.1037/a0034679>
- Harackiewicz, J. M., & Priniski, S. J. (2018). Improving student outcomes in higher education: The science of targeted intervention. *Annual Review of Psychology*, *69*, 409–435. <https://doi.org/10.1146/annurev-psych-122216-011725>
- Henning, J. A., Ballen, C. J., Molina, S., & Cotner, S. (2019). Hidden identities shape student perceptions of active learning environments. *Frontiers in Education*, *4*, 129. <https://doi.org/10.3389/feeduc.2019.00129>
- Hoffman, M., Richmond, J., Morrow, J., & Salomone, K. (2002). Investigating sense of belonging in first-year college students. *Journal of College Student Retention: Research, Theory & Practice*, *4*(3), 227–256. <https://doi.org/10.2190/DRYC-CXQ9-JQ8V-HT4V>
- Hossler, D., & Gallagher, K. S. (1987). Studying student college choice: A three phase model and the implications for policymakers. *College and University*, *62*(3), 207–221. [https://www.researchgate.net/publication/234741450\\_Studying\\_Student\\_College\\_Choice\\_A\\_Three-Phase\\_Model\\_and\\_the\\_Implications\\_for\\_Policymakers](https://www.researchgate.net/publication/234741450_Studying_Student_College_Choice_A_Three-Phase_Model_and_the_Implications_for_Policymakers)
- Jeanty, P. W. (2010). *NEARSTAT: Stata module to calculate distance-based variables and export distance matrix to text file* (Statistical Software Components S457110) [Computer program]. Boston College Department of Economics. <https://ideas.repec.org/c/boc/bocode/s457110.html>
- Johnson, S. E., & Richeson, J. A. (2009). Solo status revisited: Examining racial group differences in the self-regulatory consequences of self-presenting as a racial solo. *Journal of Experimental Social Psychology*, *45*, 1032–1035. <https://doi.org/10.1016/j.jesp.2009.04.031>
- Layous, K., Davis, E. M., Garcia, J., Purdie-Vaughns, V., Cook, J. E., & Cohen, G. L. (2017). Feeling left out, but affirmed: Protecting against the negative effects of low belonging in college. *Journal of Experimental Social Psychology*, *69*(March), 227–231. <https://doi.org/10.1016/j.jesp.2016.09.008>
- Lieberman, M. D. (2013). *Social: Why our brains are wired to connect*. Oxford University Press.
- Lord, C. G., & Saenz, D. S. (1985). Memory deficits and memory surfeits: Differential cognitive consequences of tokenism for tokens and observers. *Journal of Personality and Social Psychology*, *49*(4), 918–926. <https://doi.org/10.1037//0022-3514.49.4.918>
- Matz, R. L., Fata-Hartley, C. L., Posey, L. A., Laverty, J. T., Underwood, S. M., Carmel, J. H., Herrington, D. G., Stowe, R. L., Caballero, M. D., & Cooper, M. M. (2018). Evaluating the extent of a large-scale transformation in gateway science courses. *Science Advances*, *4*(10), eaau0554. <https://doi.org/10.1126/sciadv.aau0554>
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Mishra, S. (2020). Social networks, social capital, social support and academic success in higher education: A systematic review with a special focus on “underrepresented” students. *Educational Research Review*, *29*(February), 100307. <https://doi.org/10.1016/j.edurev.2019.100307>
- Moschetti, R. V., & Hudley, C. (2015). Social capital and academic motivation among first-generation community college students. *Community College Journal of Research and Practice*, *39*(3), 235–251. <https://doi.org/10.1080/10668926.2013.819304>
- Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*(10), 879–885. <https://doi.org/10.1111/j.1467-9280.2007.01995.x>



- Nguyen, T.-H., & Nguyen, B. M. D. (2018). Is the “first-generation student” term useful for understanding inequality? The role of intersectionality in illuminating the implications of an accepted—yet unchallenged—term. *Review of Research in Education*, 42(1), 146–176. <https://doi.org/10.3102/0091732X18759280>
- Pachankis, J. E. (2007). The psychological implications of concealing a stigma: A cognitive-affective-behavioral model. *Psychological Bulletin*, 133(2), 328–345. <https://doi.org/10.1037/0033-2909.133.2.328>
- Petty, T. (2014). Motivating first-generation students to academic success and college completion. *College Student Journal*, 48(2), 133–140. [https://go.gale.com/ps/i.do?id=GALE%7CA377286769&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=01463934&p=AONE&sw=w&userGroupName=tel\\_oweb&isGeoAuthType=true](https://go.gale.com/ps/i.do?id=GALE%7CA377286769&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=01463934&p=AONE&sw=w&userGroupName=tel_oweb&isGeoAuthType=true)
- Piff, P. K., Kraus, M. W., & Keltner, D. (2018). Unpacking the inequality paradox: The psychological roots of inequality and social class. In *Advances in Experimental Social Psychology* (Vol. 57, pp. 530124). Academic Press.
- Posse Foundation. (2020). *About the Posse Foundation*. <https://www.possefoundation.org/about-posse>
- Prentice, D. A., & Miller, D. T. (1992). When small effects are impressive. *Psychological Bulletin*, 112(1), 160–164. <https://doi.org/10.1037/0033-2909.112.1.160>
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster. <https://doi.org/10.1145/358916.361990>
- Richardson, R. C., Jr., & Skinner, E. F. (1992). Helping first-generation minority students achieve degrees. *New Directions for Community Colleges*, 1992(80), 29–43. <https://doi.org/10.1002/cc.36819928005>
- Royston, P. (2004). Multiple imputation of missing values. *Stata Journal*, 4(3), 227–241. <https://doi.org/10.1177/1536867X0400400301>
- Royston, P. (2005). Multiple imputation of missing values: Update of ice. *Stata Journal*, 5(4), 527–536. <https://doi.org/10.1177/1536867X0500500404>
- Rumberger, R. W., & Palardy, G. J. (2005). Does segregation still matter? The impact of social composition on academic achievement in high school. *Teachers College Record*, 107, 1999–2045.
- Sandefur, G. D., Meier, A. M., & Campbell, M. E. (2006). Family resources, social capital, and college attendance. *Social Science Research*, 35(2), 525–553. <https://doi.org/10.1016/j.ssresearch.2004.11.003>
- Scott, A. N., McNair, D. E., Lucas, J. C., & Land, K. M. (2017). From gatekeeper to gateway: Improving student success in an introductory biology course. *Journal of College Science Teaching*, 46(4), 93–99. [https://doi.org/10.2505/4/jcst17\\_046\\_04\\_93](https://doi.org/10.2505/4/jcst17_046_04_93)
- Sekaquaptewa, D., & Thompson, M. (2002). The differential effects of solo status on members of high-and low-status groups. *Personality and Social Psychology Bulletin*, 28(5), 694–707. <https://doi.org/10.1177/0146167202288013>
- Soria, K. M., & Stebleton, M. J. (2012). First-generation students’ academic engagement and retention. *Teaching in Higher Education*, 17(6), 673–685. <https://doi.org/10.1080/13562517.2012.666735>
- Stanton-Salazar, R. D. (1997). A social capital framework for understanding the socialization of racial minority children and youths. *Harvard Educational Review*, 67(1), 1–41. <https://doi.org/10.17763/haer.67.1.140676g74018u73k>
- Stephens, N. M., Fryberg, S. A., Markus, H. R., Johnson, C. S., & Covarrubias, R. (2012). Unseen disadvantage: How American universities’ focus on independence undermines the academic performance of first-generation college students. *Journal of Personality and Social Psychology*, 102(6), 1178–1197. <https://doi.org/10.1037/a0027143>
- Stephens, N. M., Townsend, S. S., Markus, H. R., & Phillips, L. T. (2012). A cultural mismatch: Independent cultural norms produce greater increases in cortisol and more negative emotions among first-generation college students. *Journal of Experimental Social Psychology*, 48(6), 1389–1393. <https://doi.org/10.1016/j.jesp.2012.07.008>
- Strayhorn, T. L. (2012). *College students’ sense of belonging: A key to educational success for all students*. Routledge. <https://doi.org/10.4324/9780203118924>
- Swenson, L. M., Nordstrom, A., & Hiester, M. (2008). The role of peer relationships in adjustment to college. *Journal of College Student Development*, 49(6), 551–567. <https://doi.org/10.1353/csd.0.0038>
- Tibbetts, Y., Harackiewicz, J. M., Canning, E. A., Boston, J. S., Priniski, S. J., & Hyde, J. S. (2016). Affirming independence: Exploring mechanisms underlying a values affirmation intervention for first-generation students. *Journal of Personality and Social Psychology*, 110(5), 635–659. <https://doi.org/10.1037/pspa0000049>
- Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226922461.001.0001>
- Tinto, V. (2003). Learning better together: The impact of learning communities on student success. *Higher Education Monograph Series*, 1(8), 1–8. [https://www.researchgate.net/publication/237333638\\_Learning\\_Better\\_Together\\_The\\_Impact\\_of\\_Learning\\_Communities\\_on\\_Student\\_Success](https://www.researchgate.net/publication/237333638_Learning_Better_Together_The_Impact_of_Learning_Communities_on_Student_Success)
- Turetsky, K. M., Purdie-Greenaway, V., Cook, J. E., Curley, J. P., & Cohen, G. L. (2020). A psychological intervention strengthens students’ peer social networks and promotes persistence in STEM. *Science Advances*, 6(45), eaba9221. <https://doi.org/10.1126/sciadv.aba9221>
- U.S. Census Bureau. (2017). *2017 U.S. gazetteer files: Zip code tabulation areas*. <https://www.census.gov/geo/maps-data/data/gazetteer2017.html>
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2017a). *Common Core of Data: America’s Public Schools* [Data set]. <https://nces.ed.gov/ccd/ccddata.asp>
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2017b). *Private School Universe Survey* [Data set]. <https://nces.ed.gov/surveys/pss/pssdata.asp>
- Walton, G. M., & Brady, S. T. (2020). The social-belonging intervention. In G. M. Walton & A. J. Crum (Eds.), *Handbook of wise interventions: How social-psychological insights can help solve problems* (pp. 36–62). Guilford Press.
- Walton, G. M., & Cohen, G. L. (2007). A question of belonging: race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1), 82–96. <https://doi.org/10.1037/0022-3514.92.1.82>

- Walton, G. M., & Cohen, G. L. (2011). Sharing motivation. In D. Dunning (Ed.), *Social motivation* (pp. 79–101). Psychology Press.
- Walton, G. M., Cohen, G. L., Cwir, D., & Spencer, S. J. (2012). Mere belonging: The power of social connections. *Journal of Personality and Social Psychology*, *102*(3), 513–532. <https://doi.org/10.1037/a0025731>
- Wohn, D. Y., Ellison, N. B., Khan, M. L., Fewins-Bliss, R., & Gray, R. (2013). The role of social media in shaping first-generation high school students' college aspirations: A social capital lens. *Computers & Education*, *63*(April), 424–436. <https://doi.org/10.1016/j.compedu.2013.01.004>

### Authors

KEVIN R. BINNING is an associate professor of psychology and a research scientist at the Learning Research and Development Center at the University of Pittsburgh. He studies issues related to

belonging, diversity, and equity, particularly in educational contexts.

LORRAINE R. BLATT is a doctoral candidate in the Department of Psychology at the University of Pittsburgh. She studies how public policy and structural factors influence child development and how those associations interface with socioeconomic status and race/ethnicity.

SUSIE CHEN received her PhD in psychology from the University of Pittsburgh in 2020. She is now a research scientist at WGU Labs, where she studies issues related to equity in higher education.

ELIZABETH VOTRUBA-DRZAL is a professor of psychology and a senior scientist at the Learning Research and Development Center at the University of Pittsburgh. She studies how key contexts support learning and socioemotional development during critical educational transitions.