CrossMark

Getting off to a good start? First-year undergraduate research experiences and student outcomes

Nicholas A. Bowman[1] · Joshua M. Holmes[1]

Abstract Undergraduate research is widely perceived as a "high-impact practice" that promotes students' learning, cognition, career planning, and educational attainment. With some exceptions, the existing evidence largely provides support for these beliefs. However, these studies typically examine research experiences that occur later in the undergraduate years, whereas engaging in undergraduate research during the first year is becoming increasingly common. First-year experiences may yield different outcomes than later experiences for a variety of reasons; in addition, previous studies often do not account sufficiently for self-selection into undergraduate research, which may be especially problematic for cross-sectional studies that occur in the junior or senior year. Therefore, this study examines the potential impact of first-year undergraduate research using propensity score analyses within a large, multi-institutional, longitudinal dataset. Research participation is significantly and positively related to first-year university satisfaction and fourth-year undergraduate GPA, but it is unrelated to satisfaction and grades in other years as well as graduate degree intentions, retention at the same institution, and 4-year graduation. Conditional analyses indicate that these effects are largely consistent across student demographics, pre-university achievement, and institutional selectivity.

Keywords Undergraduate research · Student success · Academic achievement · Student retention · Postsecondary students

George Kuh (2008) designated participation in undergraduate research as a "high-impact practice" and argued that students engaging in such practices are more successful on campus. Undergraduate research has been a pedagogical approach used for decades (Boyer Commission 1998; Kilgo et al. 2014; Kuh 2008; Merkel 2003). In 1998, the Boyer Commission called for the reinvention of undergraduate education centered, in particular, on research-

✉ Nicholas A. Bowman
   nick-bowman@uiowa.edu

[1]   Department of Educational Policy and Leadership Studies, N491 Lindquist Center, University of Iowa, Iowa City, IA 52242, USA

based learning. The report notes that "learning is based on discovery guided by mentoring rather than the transmission of information" (p. 24). Empirical research suggests that undergraduate research participation is linked to a wide variety of outcomes, including cognitive development, academic achievement, preparation for graduate school, professional self-efficacy, and retention and persistence (see Kilgo et al. 2014; Kuh et al. 2008; Lopatto 2006, 2010; Nagda et al. 1998).

Despite the growing body of literature on this topic, little research examines undergraduate research during students' first year in school. If undergraduate research allows students to "sink their roots in the culture of the discipline" (Merkel 2003, p. 41) as well as to explore potential career aspirations or graduate degree pursuits, then engaging in these experiences early might be beneficial. As such, many colleges and universities have begun offering first-year undergraduate research opportunities (Lopatto 2010). Moreover, as a methodological consideration, students who engage in research during their junior or senior year may have already decided to have attended graduate school (and have already persisted in postsecondary enrollment for years), so studying first-year research engagement may provide more useful and valid insights into some desired outcomes. This paper addresses the gaps in the current literature base regarding undergraduate research by using a multi-institutional, longitudinal dataset. Specifically, this study examines the following two questions:

1. What effect does first-year participation in undergraduate research have on student success outcomes (i.e., undergraduate GPA, university satisfaction, intentions for graduate school, retention, and 4-year graduation)?
2. Do these effects differ by student demographic characteristics (i.e., race/ethnicity, sex, parental education, and standardized test scores) and institutional selectivity?

By accounting for self-selection with propensity score analyses that includes a variety of pre-university and institutional characteristics, this study can draw stronger inferences about the potential causal effects of research participation in during the first year. In addition, the consideration of conditional effects provides important and novel insights into for whom this engagement may be most beneficial.

Undergraduate research experiences and outcomes

The goal of many undergraduate research opportunities is "to involve students with actively contested questions, empirical observation, cutting-edge technologies, and the sense of excitement that comes from working to answer important questions" (Kuh 2008, p. 10). Kuh attributes these changes to be a function of collaborative work with faculty. He suggests that undergraduate research experiences provide students with an individual, deep connection with a faculty member or graduate student. This connection allows for individualized, prompt, and continuous feedback during the research collaboration. Additionally, this experience provides exposure to challenges in research and ways in which one can overcome these difficulties. Undergraduate research experience may provide benefits that warrant its label as a high-impact practice, including by promoting student persistence and degree completion.

Proponents of undergraduate research initially argued that the effect of undergraduate research participation would be most pronounced on students' advanced degree intentions (Boyer Commission 1998). In other words, as a pedagogical practice, undergraduate research

helps students clarify educational objectives and plans, in particular their postgraduate degree aspirations. Zydney et al. (2002) found that students participating in formal undergraduate research programs self-reported significantly greater gains in academic degree goals than those who did not participate in undergraduate research, and students involved with informal research opportunities had significantly higher self-reported gains in academic career goals than their peers who did not complete any research opportunity. Similarly, Kim and Sax (2009) found greater postgraduate degree plans for students who participate in research experiences with faculty than those who do not. A rigorous examination by Kilgo and Pascarella (2016) also found that undergraduate research is associated with greater graduate degree intentions even when controlling for a variety of pre-university and institutional attributes (albeit with no significant effect on 4-year graduation). This relationship was across a variety of student characteristics, including sex and race.

Additional studies provide nuance and insights into the nature of these degree and career aspirations. Seymour et al. (2004) interviewed 76 students who participated in undergraduate research three times (twice during university and again 20 months after graduation). They also collected data on a comparison group of 63 students. The researchers' findings suggest no support for the assertion that "undergraduate research experiences had prompted rising seniors to choose particular careers" (p. 530), but rather, they found that the experience "had clarified, refined, or confirmed students' pre-existing choice of career directions" (p. 530). The vast majority of students described the experience of undergraduate research as "an educational and personal-growth experience with many transferable benefits" (p. 530). In other words, the positive career related benefits resulting from undergraduate research may come from increased student self-efficacy and a stronger individual connection with a faculty member. Adedokun et al. (2013) also argue that research self-efficacy largely explained the link between undergraduate research participation and students' advanced degree intentions.

Kuh (2008) asserts that undergraduate research (and other high-impact practices) are more beneficial for underrepresented students than for majority students. As a pedagogical practice, universities specifically have promoted undergraduate research opportunities to increase underrepresented students' participation and enculturation into STEM major fields. Kim and Sax (2009) found that research-related student-faculty interactions were more strongly associated with grades for African-American students than for Latino and Asian-American students. Studies focusing on undergraduate research experiences suggest that the experience conditionally affects retention; specifically, program participation was positively related to retention only for African-American students, not for white or Latino/Hispanic students (Gregerman et al. 1998). These findings also suggest that the program may be most beneficial for African-American students who were below median academic achievement for their racial group. This study also disaggregated the time students chose to engage in undergraduate research, finding stronger relationships with retention for students who first participate in their sophomore year (rather than other years). Hurtado et al. (2008) found similar gains in science identities and self-efficacy for first-year, African-American students associated with structured undergraduate research programs in science fields. It may be that these benefits occurred for students of color because of a structured first-year opportunity to engage in research as well as a strong peer network.

Prompted by this earlier research suggesting positive associations between undergraduate research participation and retention for underrepresented students, Jones et al. (2010) longitudinally explored the effect of undergraduate research participation among biology majors. Participation in an undergraduate research experience, regardless of timing, leads to a greater

likelihood of graduation, attaining a degree in biology, and receiving a cumulative GPA of 3.0 or higher. Their findings suggest that participation in undergraduate research experiences is most valuable for underrepresented student populations, especially in terms of graduating with any degree (although well over 90% of undergraduate researchers from any racial/ethnic group received a degree). Similarly, Kim and Conrad (2006) observed that historically black colleges and universities have higher rates of undergraduate research experience for African-American students than do predominantly white institutions. Finally, Hathaway et al. (2002) found that underrepresented students participating in a formal undergraduate research program were more likely to pursue postgraduate degrees than their peers in non-structured programs or with no research experience at all. For underrepresented students, participation in undergraduate research programs may provide socialization into the discipline and dispel myths about graduate school.

However, the literature has some noteworthy limitations. First, and perhaps most importantly, many of these studies confound research participation and retention, since students who have dropped out of their college or university are no longer able to participate in undergraduate research. Thus, the reverse causal direction may be quite likely: Postsecondary attrition leads to a lack of engagement with research, instead of research experiences leading to subsequent retention, graduation, and pursuit of a graduate degree. Seymour et al. (2004) also illustrated this potential directionality problem even among students who persist, since research experiences clarified and reinforced students' pre-existing career and educational goals rather than creating those goals and intentions in the first place. Second, in a related concern, even studies that avoid this problem through careful design do not sufficiently account for selection bias into these research opportunities. That is, students who decide to participate in undergraduate research may differ considerably in important ways from those who do not (e.g., Kilgo and Pascarella 2016), so properly adjusting for these differences is crucial for drawing strong conclusions about the impact of this practice. Third, a majority of the studies are often limited to small sample sizes or samples from single institutions, so generalizability may be limited.

Theoretical framework

This paper draws upon two closely related and highly influential frameworks: Pace's (1982) quality of effort and Astin's (1984) theory of involvement. Both theories are complimentary, in that they assert that the more students put into their undergraduate experiences, the more they will benefit from them. Pace (1982) argues that the quality of students' effort largely dictates their learning and success outcomes; moreover, high-quality effort should lead to greater academic achievement, which in turn promotes satisfaction, retention, and graduation. Astin's theory of involvement expands Pace's argument to include the quantity of engagement as well. Together, these two frameworks suggest the importance of both the depth and the breadth of experiences for shaping a variety of short-term and long-term outcomes.

Undergraduate research, particularly during students' first year, is notable for the quality and quantity of effort required. The actual amount of time that students spend depends upon a variety of factors, but 6–10 hours per week is typical (e.g., Kuh et al. 2007; Zydney et al. 2002). Given that many students do not spend more than 10 hours per week doing academic work outside of class (Arum and Roksa 2011), this quantity of engagement is notable. Moreover, the intellectual demands for understanding and conducting research during the first

year—before most students have had much exposure to research methodology or disciplinary content coursework—are substantial. Thus, the quality of effort required is likely to bolster students' subsequent achievement, satisfaction, retention, and degree intentions.

Present study

The present study sought to overcome limitations in the previous literature. Specifically, we used a multi-institutional, longitudinal dataset to explore the effect of first-year participation in research experiences on undergraduate GPA, university satisfaction, intention to obtain a graduate degree, retention, and 4-year graduation. This study also provides stronger claims about potential causality through two methodological decisions. First, propensity score analyses accounted for self-selection into undergraduate research. Within the analysis, the large scale of this data collection and longitudinal design of this dataset allowed us to consider a variety of potentially confounding variables that occurred before beginning postsecondary studies and that extend well beyond usual demographics and pre-university achievement. Specifically, these covariates also included intended undergraduate major, highest intended degree, academic dispositions, various academic and social experiences in the previous year, and institutional characteristics.

Second, this study focused on undergraduate research that occurs during the first year. In the junior and senior year, students have already likely made decisions about their future career goals, so undergraduate research might clarify—rather than increase—decisions to attend graduate school (Seymour et al. 2004). Students who engage in this experience late in their undergraduate careers are probably doing so *because* they are high achieving and seeking to prepare for graduate school. As a result, the direction of causality may be the opposite of what some studies—especially those with cross-sectional designs—might suggest. Measuring experiences in the first year with a longitudinal dataset also enables us to predict outcomes that occur years after the experience. That said, first-year undergraduate research might simply have different effects than later research participation, which occurs when students have taken more coursework within their field of study.

Finally, this study investigated whether and how first-year undergraduate research varies across several groups, including students' race/ethnicity, sex, parental education, and standardized test scores as well as institutional selectivity. To date, very little is known about such variations, in part because the size and multi-institutional nature of this dataset are uncommon in research on this topic.

Method

Data source and participants

This study used data from the Wabash National Study of Liberal Arts Education. Colleges and universities were selected to participate based on their strong commitment to liberal arts education. The sample contained 46 4-year institutions, which included religiously affiliated, single-sex, and minority-serving schools. Institutions exhibited a wide range of selectivity, tuition costs, and geographic diversity. Given our interest in examining outcomes of

undergraduate research up to several years after postsecondary entry, the three 2-year institutions within this dataset were excluded from the analytic sample.

Students beginning their first year were invited to participate in a longitudinal study. Before classes began or during their first 2–3 weeks on campus (time 1), 16,719 students at 4-year institutions completed a registration form that included demographic information; a questionnaire of various high school experiences, interests, attitudes, and values; and a battery of assessments. Approximately 2 weeks before the end of their first year (time 2), students who took part in the initial assessment were invited to participate in a second round of data collection. They completed the same battery of assessments, along with questionnaires that asked about their university experiences, interests, attitudes, and values. A total of 8475 students participated in the second wave, yielding a retest response rate of 51%. A third wave of surveys was administered at the end of students' fourth year; the questionnaires and assessments were essentially identical to those used in wave 2. Of students at 4-year institutions who participated in the first two waves, 4211 also participated in the third wave, which constitutes a retest response rate of 50%.

To provide some adjustment for potential non-response bias, a sample weighting algorithm was developed and implemented to make the sample more representative of the incoming first-year classes of those institutions in terms of sex, race, academic ability, and institutional type. In many surveys of college students and other adults, women and whites tend to be somewhat overrepresented within the sample (Groves et al. 2009); these groups are also more likely to persist within college (Radford et al. 2010) and therefore persist within the longitudinal study. The weighting strategy adjusted for differences between the target population (i.e., the incoming cohort of undergraduates at these colleges and universities) and study participants that resulted from unit non-response (see Biemer and Christ 2008). The weights were normalized with a mean of 1.0 so that weighting did not affect the analytic sample size. Overall, 56.6% of participants were female, 10.4% were black/African-American, 5.6% were Asian-American/Pacific Islander, 4.8% were Latino/Hispanic, and 2.5% were from another race/ethnicity. Moreover, 5.2% of participants engaged in undergraduate research during their first year; this figure is similar to results from the National Survey of Student Engagement (2016).

Measures

Dependent variables Undergraduate GPA was measured via student self-reports at the end of the first year and the end of the fourth year on an eight-point scale (1 = C− or lower, to 8 = A). University satisfaction was computed as the index of two items that were assessed at those same two timepoints: "How would you evaluate your entire educational experience at this institution?" (1 = poor, to 4 = excellent) and "If you could start over again, would you go to the *same institution* you are now attending?" (1 = definitely no to 4 = definitely yes; Cronbach's alpha = 0.70 in the first year and 0.71 in the fourth year). These academic achievement and satisfaction measures were subsequently standardized with a mean of zero and a standard deviation of one, allowing the results of the propensity score analyses to be interpreted as Cohen's ds (i.e., the standardized mean difference) between students who did and did not participate in undergraduate research. Students' intentions to receive a graduate degree were indicated during the first and fourth years with a binary variable (0 = no, 1 = yes). Three university retention variables indicated whether students were enrolled at their initial institution in the Fall term of the second year, third year, and fourth year. On-time graduation

was indicated by whether the student had graduated from that institution by the end of the fourth year.

Independent variables The primary independent variable was participation in undergraduate research during the first year (0 = no, 1 = yes). As discussed earlier, this first-year measure was used to explore outcomes that might result from research participation later in students' undergraduate years.

As recommended for this analytic technique, all variables included in the propensity score were pre-university characteristics (i.e., before students participated in the treatment) that were selected based on their expected impact on participation in the treatment and/or outcome variables (see Austin 2011; Guo and Fraser 2015). Several different categories of predictors—with multiple variables from each category—were used to create the propensity score; this approach is often particularly effective in substantially reducing or eliminating bias (Steiner et al. 2015). This study included three important categories of predictors that are essential for predicting future behavior but are often not used as covariates in postsecondary research. First, students' pre-university dispositions and traits may predict both undergraduate research participation and postsecondary success. Academic motivation reflects interest in engaging with academic experiences and course material (8 items, $\alpha = 0.69$). Need for cognition indicates a preference to engage in cognitively challenging and thought-provoking activities (18 items, $\alpha = 0.89$; Cacioppo et al. 1996). Students also rated the personal importance of various life goals, and three of these indices were used: professional success (5 items, $\alpha = 0.76$), contributing to science (2 items, $\alpha = 0.70$), and political and social involvement (11 items, $\alpha = 0.80$).

Second, postsecondary intentions were assessed for highest intended degree (1 = vocational/technical certificate or diploma to 6 = doctorate degree) and undergraduate major (dummy-coded variables for allied health, business, education, engineering, humanities/fine arts, mathematics/statistics, natural sciences, and others, with social sciences as the referent group). Third, the analyses included various forms of academic and social engagement during the high school (or secondary school) senior year, because these forms of participation likely predict postsecondary academic involvement. The pre-university experiences were assessed using a five-point scale (1 = never, to 5 = very often): studying alone, studying with friends, talking with teachers, volunteering, and working for pay. Standardized test scores were indicated through the ACT composite score or the converted SAT math plus verbal scores. High school GPA was indicated with two dummy-coded variables for an average of B (B+ to B−) and an average of C+ or lower, with A− to A+ as the referent group.

Demographics were used, since they frequently predict postsecondary student engagement and success outcomes (e.g., Kinzie et al. 2007; Radford et al. 2010). These variables included race/ethnicity (dummy-coded variables for Asian-American/Pacific Islander, black/African-American, Latino/Hispanic/Chicano, and other race/ethnicity, with white/Caucasian as the referent group), sex (0 = female, 1 = male), and parental education (the average of mother's and father's level of education). Finally, given the multi-institutional nature of this study, institutional characteristics were also included. Institutional type was measured with dummy-coded variables for regional university and research university, with liberal arts college as the referent group. Institutional selectivity was indicated with the Barron's index (1 = non-competitive to 6 = most competitive). Many of these measures have been used in previous research and have strong evidence for their content, construct, and predictive validity; detailed information is available from the Center of Inquiry at

Wabash College (2016) and the Center for Research on Undergraduate Education at the University of Iowa (2008).

Analyses

Two key considerations guided the choice of the matching approach for these propensity score analyses. First, because the current sample contained students nested within institutions, hierarchical linear modeling (HLM) analyses were used. This nesting violates a key assumption of ordinary least squares (OLS) multiple regression, whereas HLM accounts for this issue by partitioning the variance within institutions (at level 1) and between institutions (at level 2) and adjusting standard errors accordingly (e.g., Heck and Thomas 2009; Raudenbush and Bryk 2002). According to the intraclass correlation coefficients (ICCs), substantial between-institution variance was apparent for all outcomes, including university satisfaction (ICC = 0.090 in the first year and 0.093 in the fourth year), undergraduate GPA (ICC = 0.094 and 0.098, respectively), intent to receive a graduate degree (ICC = 0.094 and 0.067, respectively), retention (ICC = 0.126 to year 2, 0.146 to year 3, and 0.177 to year 4), and 4-year graduation (ICC = 0.361).

Second, only a small proportion of students participated in undergraduate research during the first year, so a strategy that pairs each research participant with only one student who did not participate would result in a substantial loss of statistical power. Therefore, given the multilevel structure of the data and the desire to retain as many students within the sample as possible, stratification was used to conduct matching (for examples of multilevel propensity score stratification, see Bowman et al. 2016; Hong and Raudenbush 2006). Stratification is one of several techniques that compare the outcomes of students in the treatment and control conditions who are very similar in their predisposition to engage in the treatment (Austin 2011; Guo and Fraser 2015). The current study used stratification to match students both within and across institutions while also accounting for the multilevel structure of the data; such approaches provide better results than ignoring the multilevel structure and/or attempting to match solely within the same level 2 unit (Vaughan et al. 2014; Wang 2015). Power analyses using Power Up! Software (Dong and Maynard 2013) found that this design has considerable statistical power, since it would be able to detect an effect of $d = 0.14$ with power of 0.80; this effect size is slightly below the value designated for "small" effects within social science research generally (Cohen 1988) and postsecondary impact research specifically (Mayhew et al. 2016). Therefore, this study has ample power to identify meaningful effects if they exist.

To help select variables for the propensity score, each pre-university variable was entered at the appropriate level (student or institution) as the lone predictor of undergraduate research participation in a multilevel analysis. Based on results from simulation studies (Brookhart et al. 2006; Patrick et al. 2011), some variables were retained for creating the propensity score if they were related to the outcome but did not significantly predict undergraduate research participation; these included demographics, high school grades, time spent studying, working for pay, and institutional type. In contrast, a few variables that did not significantly predict research participation—and for which the existing literature does not provide evidence for a link with this study's outcomes—were excluded from the propensity score (high school extracurricular activities, high school socializing, importance of contributing to the arts, and eudaimonic well-being).

The logit was used to compute a single, linear propensity score. Figure 1 provides the distribution of this propensity score for students who did and did not participate in
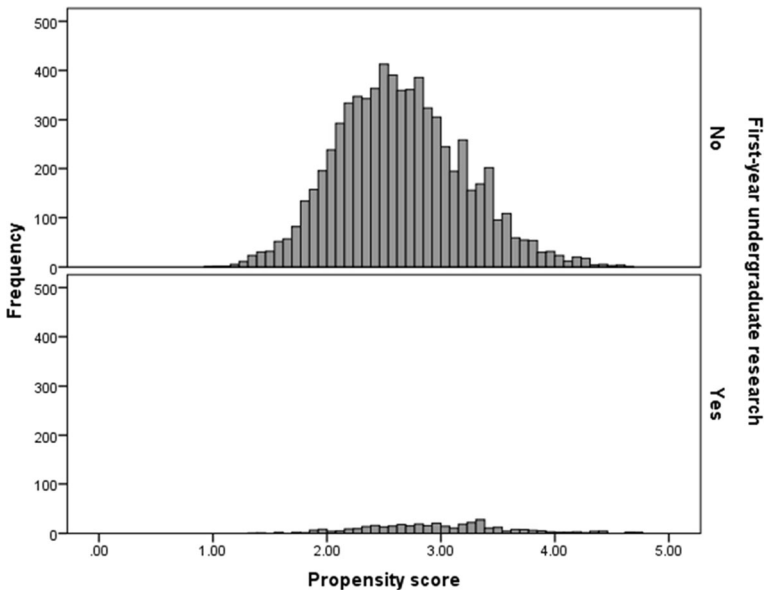
Fig. 1 Histogram of propensity scores for students who did and did not participate in first-year undergraduate research

undergraduate research. As a first step, the linear propensity score variable was divided into five equal strata with 20% of participants included per stratum to equate participants on the propensity score within each stratum (Cochran 1968). This approach failed to yield sufficient balancing, so greater numbers of strata were tested. The final solution employed 30 strata, and it removed participants from the highest stratum, which was still unbalanced, from the analytic sample. After doing so, a two-way analysis of variance predicting the linear propensity score with strata and treatment condition (research vs. no research) as independent variables found no significant main effect of treatment condition and no interaction between strata and treatment. This large number of strata means that students in the treatment and control conditions within the same stratum had very similar propensities to participate, and the large sample size of this study permits the use of this many strata while still providing ample statistical power.

Another test of the effectiveness of the propensity score balancing examines whether each variable used to create the propensity score significantly predicts program participation when including the propensity score adjustment. If the propensity score succeeds in removing self-selection bias, then the pre-university variable should not significantly predict undergraduate research participation when performing the adjustment. Table 1 provides a summary of these multilevel tests; none of the independent variables significantly predicted research participation when the PSM adjustment occurred, which indicates that the propensity score successfully removed bias associated with those variables.

The final analyses predicted each of the ten student success outcomes with undergraduate research participation and the PSM strata entered as predictors. University satisfaction and GPA were treated as continuous outcomes through HLM analyses. Hierarchical generalized linear modeling (HGLM) analyses were used to predict the dichotomous outcomes of graduate degree intentions, retention, and graduation. To explore whether the potential impact of undergraduate

Table 1 Hierarchical generalized linear modeling analyses predicting first-year undergraduate research participation before and after propensity score balancing

Predictor	Before balancing			After balancing		
	B	SE	p	B	SE	p
Pre-university dispositions and traits						
Academic motivation	0.389	0.129	0.003	0.020	0.133	.880
Need for cognition	0.355	0.091	<0.001	− 0.051	0.101	.610
Importance of professional success	0.384	0.083	<0.001	0.031	0.095	.747
Importance of contributing to science	0.393	0.072	<0.001	0.053	0.068	.434
Importance of social/political involvement	0.352	0.129	0.006	− 0.040	0.077	.601
Physical health	0.262	0.095	0.006	0.070	0.104	.502
Postsecondary intentions						
Intended allied health major	− 0.349	0.170	0.040	0.051	0.190	.788
Intended business major	− 0.094	0.108	0.386	− 0.042	0.101	.679
Intended education major	− 0.281	0.199	0.159	− 0.003	0.236	.991
Intended engineering major	0.141	0.257	0.583	0.074	0.263	.780
Intended humanities or fine arts major	− 0.421	0.197	0.033	− 0.039	0.226	.864
Intended mathematics or statistics major	0.115	0.436	0.792	− 0.172	0.618	.781
Intended natural sciences major	0.723	0.136	<0.001	0.061	0.212	.774
Intended other major	0.077	0.140	0.582	− 0.136	0.150	.363
Highest intended degree	0.075	0.047	0.108	0.025	0.044	.568
High school engagement						
High school teacher interactions	0.219	0.053	<0.001	− 0.062	0.065	.335
High school studying with friends	0.125	0.068	0.069	− 0.053	0.094	.574
High school studying alone	− 0.003	0.082	0.968	0.013	0.067	.849
High school volunteering	0.145	0.053	0.006	− 0.055	0.076	.468
High school working for pay	0.076	0.070	0.273	− 0.005	0.067	.945
High school achievement						
Standardized test scores	− 0.039	0.019	0.040	− 0.005	0.024	.848
B HSGPA	0.017	0.126	0.893	0.038	0.141	.787
C or lower HSGPA	0.198	0.226	0.380	0.170	0.331	.607
Demographics						
Asian-American/Pacific Islander	− 0.141	0.398	0.723	− 0.063	0.417	.881
Black/African-American	0.230	0.306	0.453	0.061	0.218	.779
Latino/Hispanic/Chicano	0.202	0.288	0.484	− 0.108	0.378	.774
Other race/ethnicity	− 0.254	0.316	0.422	0.006	0.320	.986
Male	0.128	0.128	0.315	− 0.014	0.119	.904
Parental education	− 0.029	0.033	0.384	− 0.000	0.037	.997
Institutional characteristics						
Regional university	0.343	0.292	0.246	0.033	0.241	.891
Research university	− 0.196	0.388	0.616	− 0.048	0.359	.895
Institutional selectivity	− 0.199	0.106	0.066	− 0.012	0.097	.901

Note. All variables used to create the propensity score were assessed upon entering university. Variables that did not significantly predict the treatment were still used within the propensity score if previous literature showed that they predicted the outcomes

research varies across groups, additional analyses included interaction terms between research participation and several student-level variables (race/ethnicity, sex, parental education, and standardized test scores). To increase statistical power, students of color were combined into a single group for this analysis and the corresponding interaction. Furthermore, a cross-level interaction between institutional selectivity and undergraduate research was used. Variables for research participation, the relevant moderator, and the interaction term were entered simultaneously into the equation to model these interactions appropriately (Jaccard and Turrisi 2003). Separate analyses examined each interaction to avoid problems with multicollinearity.

Limitations

Some limitations should be noted. First, although quasi-experimental designs can yield results that better estimate causal effects, these analyses only match participants on the variables used to create the propensity score. Therefore, it is possible that this study has not completely eliminated selection bias, since other relevant variables may not have been included (e.g., a direct measure of students' interests in research). Second, this study measured participation in undergraduate research during students' first year, but this variable does not specify how long students participated or what types of experiences they had. As a result, the findings cannot provide insight into implications for the length and content of such initiatives. Third, the retention and graduation indicators only reflect outcomes at students' initial college or university. Undergraduate research participation may also influence whether students persist or receive a degree from any institution; however, since retention and graduation outcomes were specific to the original institutions, persistence and graduation from other institutions could not be assessed.

Results

Table 2 displays the results for undergraduate research participation and student outcomes before and after the propensity score adjustment. Although the coefficients differ somewhat depending on whether these quasi-experimental analyses are employed, the pattern of significant results is identical. Specifically, undergraduate research is positively and significantly related to fourth-year undergraduate GPA and first-year university satisfaction. However, the results are non-significant in all other analyses, including GPA and satisfaction in the other years, 4-year graduation, and retention and graduate degree intentions in any year. Among the non-significant results, half of the coefficients were positive, and half were negative.

As shown in Table 3, additional analyses examined interactions between undergraduate research and key predictors at the student and institutional levels. Undergraduate research participation is

Table 2 Results of hierarchical linear modeling analyses for first-year undergraduate research participation predicting postsecondary student success outcomes

Outcome variable	No PSM adjustment		PSM adjustment	
	B	SE	B	SE
Undergraduate GPA in the first year	0.012	0.094	0.022	0.101
Undergraduate GPA in the fourth year	0.172*	0.075	0.165**	0.060
University satisfaction at end of first year	0.174**	0.055	0.176***	0.039
University satisfaction at end of fourth year	0.099	0.076	0.081	0.092
Intent to obtain graduate degree during the first year	− 0.040	0.172	− 0.001	0.208
Intent to obtain graduate degree during the fourth year	− 0.092	0.437	− 0.276	0.449
Retention to fall of second year	− 0.287	0.237	− 0.087	0.213
Retention to fall of third year	0.006	0.165	0.008	0.182
Retention to fall of fourth year	− 0.152	0.147	− 0.135	0.203
Graduated within 4 years	0.059	0.131	0.099	0.152

Note. Undergraduate GPA and university satisfaction were standardized with a mean of zero and a standard deviation of one. Degree intentions, retention, and graduation were modeled as binary outcomes using hierarchical generalized linear modeling analyses

*$p < .05$; **$p < .01$; ***$p < .001$

Table 3 Results of interactions between first-year undergraduate research and key variables within propensity score analyses predicting postsecondary student success outcomes

Outcome variable	Institutional selectivity	ACT/SAT score	Parental education	Student of color	Male student
Undergraduate GPA in the first year	0.150* (0.064)	0.024* (0.010)	0.074* (0.034)	− 0.017 (0.150)	0.277 (0.168)
Undergraduate GPA in the fourth year	0.112* (0.055)	− 0.003 (0.020)	0.037 (0.043)	− 0.102 (0.261)	0.267 (0.163)
University satisfaction at end of first year	− 0.041 (0.044)	− 0.008 (0.011)	0.061 (0.033)	0.118 (0.107)	0.011 (0.143)
University satisfaction at end of fourth year	− 0.049 (0.076)	− 0.016 (0.019)	− 0.019 (0.035)	0.449* (0.182)	0.172 (0.122)
Intent to obtain graduate degree during the first year	0.231 (0.163)	0.063* (0.030)	0.038 (0.090)	0.220 (0.368)	0.966*** (0.287)
Intent to obtain graduate degree during the fourth year	0.241 (0.284)	− 0.064 (0.052)	− 0.080 (0.098)	1.558 (0.973)	− 1.182* (0.570)
Retention to fall of second year	0.047 (0.178)	0.047 (0.065)	0.320** (0.115)	− 0.293 (0.467)	0.192 (0.376)
Retention to fall of third year	− 0.101 (0.139)	0.008 (0.046)	0.150 (0.098)	− 0.161 (0.344)	− 0.059 (0.339)
Retention to fall of fourth year	− 0.335* (0.140)	− 0.041 (0.055)	0.016 (0.100)	0.204 (0.463)	− 0.068 (0.396)
Graduated within 4 years	− 0.002 (0.125)	0.000 (0.032)	− 0.015 (0.056)	− 0.376 (0.406)	− 0.432 (0.286)

Note. Standard errors are in parentheses. Undergraduate GPA and university satisfaction were examined with hierarchical linear modeling analyses; these outcomes were standardized with a mean of zero and a standard deviation of one. Degree intentions, retention, and graduation were examined with hierarchical generalized linear modeling analyses to predict these binary outcomes. First-year undergraduate research participation and the corresponding predictor were included as additional variables in all analyses (with demographics modeled at level 1 and institutional selectivity at level 2)

$*p < .05$; $**p < .01$; $***p < .001$

more positively related to first-year GPA among students who attended more selective institutions, had stronger standardized test scores, and had greater parental education. Aside from these results, the other conditional effects are often non-significant and are mixed across interaction terms and outcomes. Undergraduate research is more positively related to graduate degree intentions during the first year among students with higher test scores and among male students, but this link is more positive for female students when predicting graduate degree intentions during the fourth year. In addition, the impact of undergraduate research on GPA in the fourth year is more positive at selective institutions, whereas the impact on retention to the fourth year is actually more negative at selective schools. Finally, the relationship with retention to the second year is more positive among students with high parental education.

Discussion and conclusion

Overall, first-year undergraduate research participation is positively related to fourth-year undergraduate GPA as well as first-year university satisfaction. This study's methodological and analytic approach adds credence to the argument within the current educational attainment literature suggesting positive benefits from participation in research experiences (e.g., Adedokun et al. 2013; Jones et al. 2010). Engaging in research experiences likely facilitates more frequent student-faculty interaction and possible mentoring relationships, which may be the reason that participating students are more satisfied. Interestingly, first-year participation in this experience does not have a significant effect on students' first-year GPA, but rather an apparent delayed effect on fourth-year GPA. It is possible that cognitive skills learned during research participation in the first year benefit students in more advanced coursework, which is more cognitively demanding. Previous literature has shown gains in critical thinking, writing, and communication skills as a result of research participation that are often related to higher-level coursework (Adedokun et al. 2013; Lopatto 2006, 2010). These effects on grades and satisfaction are considered small according to recommendations for higher education research (Mayhew et al. 2016) and social science more generally (Cohen 1988). However, as Cohen argues, a small effect size does not imply a lack of practical significance, because the treatment may be implemented with varying degrees of effectiveness, and the key variables will contain some degree of error. Most postsecondary experiences have a modest effect on student outcomes (Mayhew et al. 2016), so these findings for undergraduate research are on par with other forms of engagement.

In contrast, research participation was not significantly related to graduate degree intentions, retention at the same institution, or 4-year graduation. As discussed earlier, this study had sufficient statistical power to identify even small effect sizes ($d = 0.14$), so this study would likely have identified any practically meaningful effects if they were present. The mix of positive and negative coefficients for the non-significant results further supports the lack of meaningful effects for most outcomes, whereas previous studies have generally shown positive relationships associated with research experiences and similar outcomes (Kilgo and Pascarella 2016; Kim and Sax 2009; Zydney et al. 2002). At least one of two explanations could account for this divergence. First, the first-year research experiences examined in this study may simply have differential outcomes than later experiences (which were the focus of previous inquiry). In later years, students could be more developmentally prepared for research, have more relevant knowledge and skills, engage in more meaningful and autonomous activities, and/or work more collaboratively with faculty (rather than with graduate assistants or by

themselves). Any of these explanations could lead to better outcomes when undergraduate research occurs later. Second, this study conducted a rigorous examination of causal effects by using propensity score analyses, incorporating a rigorous set of covariates, examining a large dataset of diverse students and institutions, conducting longitudinal analyses that ensured the experience occurred before the outcomes, and assessing a later experience after students have already made a decision to attend graduate school. Some previous studies may have overestimated the effects of undergraduate research because of their sampling, variables, analyses, or overall research design.

The conditional findings for first-year undergraduate research participation were mixed across the various student success outcomes. Among groups that are more academically prepared or privileged (i.e., students attending more selective institutions, students with higher standardized test scores, and students with greater parental education), participation in undergraduate research in the first year was associated with more positive results. The potential effects on undergraduate GPA in the first and fourth years are more positive at selective institutions, whereas the effect on fourth-year retention is more negative at selective institutions. The one consistent pattern is that undergraduate research has a stronger effect on first-year GPA among groups that are more academically prepared or whose parents have higher levels of education. Perhaps surprisingly, these groups are not more likely to engage in first-year research; in fact, students' test scores are actually inversely related to participation (as shown in Table 1). The reason for this pattern is unclear. Perhaps students with greater academic preparation are better able to learn from and excel within their research experiences, which have the potential to be particularly challenging for first-year students. Or perhaps, these students are already intending on participating in these experiences, but at a later date, using their first year as a transition period. These students from privileged backgrounds may also have greater cultural capital that helps them know what to expect from working with faculty and possibly also tailor these experiences to fit better with their own interests and skills.

The only significant finding by race/ethnicity is that the effect of undergraduate research is greater for university satisfaction at the end of their fourth year among students of color than among white students. The fact that this interaction is significant at the end of the fourth year, but not the end of the first year, is interesting. It may be that the experience of first-year research creates different paths for these students to have different experiences in the subsequent years that might increase their satisfaction, or this first-year experience increases a sense of purpose or belonging among students of color that manifests later as satisfaction with their collegiate experience. This study did not uncover any other significant findings by race/ethnicity, contrary to other researchers' findings regarding GPA and graduation (see Jones et al. 2010; Kim and Conrad 2006; Kim and Sax 2009). The lack of additional significant findings, contrary to prior research, could be attributable to the more robust and rigorous research design.

When viewing the conditional effect of first-year participation in undergraduate research by sex, intentions to receive a graduate degree are the only significant outcome. First-year research participation is more positively related to graduate degree intentions among male students (than among female students) when this outcome is assessed during the first year, but the exact opposite is true when the outcome is assessed during the fourth year (i.e., the effect is significantly more positive for female than for male students). Participation in the first year might help female students develop research self-efficacy (Adedokun et al. 2013) that could later lead to decisions to pursue a graduate degree. For female students, this experience might help develop or clarify a STEM mindset in which female students make a decision in subsequent years to pursue advanced degrees.

Further, for male students, the research experience in the first year could work in the opposite direction, leading male students to make career choices that differ from original plans. Another explanation might lie in the relatively large number of interactions tested. These divergent results may be stem from one or more significant relationships that simply occurred by random chance. For instance, the greater effect for male students in the first year is highly significant, whereas the greater effect for female students in the fourth year is close to the cutoff of $p < .05$. Therefore, this latter result may not be substantively meaningful, especially given that significance testing is expected to yield an occasional type I error within a fairly large number of tests. Further studies conditionally exploring first year undergraduate research experiences by sex could help clarify these results.

In summary, this study finds that students participating in first-year undergraduate research experiences have higher GPAs in their fourth year and are more satisfied with their postsecondary experience in their first year. The results suggest that colleges and universities should consider directing resources in encouraging this type of experience for first-year students, while weighing the fact that this practice does not seem as effective at promoting student success as some may believe. Further research is needed on the overall effects of first-year undergraduate research to better understand these dynamics, along with how these interact with institutional selectivity and student characteristics. In particular, are there certain forms of early research engagement that are more effective at promoting student outcomes than others? Is research conducted with faculty during intensive summer experiences more effective? And are these experiences more beneficial for some groups than for others?

References

Adedokun, O. A., Bessenbacher, A. B., Parker, L. C., Kirkham, L. L., & Burgess, W. D. (2013). Research skills and STEM undergraduate research students' aspirations for research careers: mediating effects of research self-efficacy. *Journal of Research in Science Teaching, 50*(8), 940–951.

Arum, R., & Roksa, J. (2011). *Academically adrift: limited learning on college campuses*. Chicago: University of Chicago Press.

Astin, A. W. (1984). Student involvement: a developmental theory for higher education. *Journal of College Student Personnel, 25*(4), 297–308.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*, 399–424.

Biemer, P. P., & Christ, S. L. (2008). Weighting survey data. In E. D. de Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 317–341). New York: Psychology Press.

Bowman, N. A., Denson, N., & Park, J. J. (2016). Racial/cultural awareness workshops and post-college civic engagement: a propensity score matching approach. *American Educational Research Journal, 53*(6), 1556–1587.

Boyer Commission on Educating Undergraduates in the Research University. (1998). *Reinventing undergraduate education: a blueprint for America's research universities*. Stony Brook: State University of New York.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology, 163*(12), 1149–1156.

Cacioppo, J., Petty, R., Feinstein, J., & Jarvis, W. (1996). Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*, 197–253.

Center for Research on Undergraduate Education. (2008). *Quantitative methods section for the Wabash National Study of Liberal Arts Education*. Iowa City: Author, University of Iowa Retrieved from https://education.uiowa.edu/sites/education.uiowa.edu/files/documents/centers/crue/research_methods_draft_march2008.pdf.

Center of Inquiry. (2016). *Outcomes and experiences measures: Wabash National Study 2006–2012*. Crawfordsville: Author, Wabash College Retrieved from http://www.liberalarts.wabash.edu/study-instruments/.

Cochran, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics, 24*, 295–313.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah: Lawrence Erlbaum.

Dong, N., & Maynard, R. A. (2013). *Power Up!* A tool for calculating minimum detectable effect sizes and sample size requirements for experimental and quasi-experimental designs. *Journal of Research on Educational Effectiveness, 6*, 24–67.

Gregerman, S. R., Lerner, J. S., von Hippel, W., Jonides, J., & Nagda, B. A. (1998). Undergraduate student-faculty research partnerships affect student retention. *The Review of Higher Education, 22*(1), 55–72.

Groves, R. M., Fowler Jr., F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken: Wiley.

Guo, S., & Fraser, M. W. (2015). *Propensity score analysis: statistical methods and applications* (2nd ed.). Los Angeles: Sage.

Hathaway, R. S., Nagda, B. A., & Gregerman, S. R. (2002). The relationship of undergraduate research participation to graduate and professional education pursuit: an empirical study. *Journal of College Student Development, 43*(5), 614–631.

Heck, R. H., & Thomas, S. L. (2009). *An introduction to multilevel modeling techniques* (2nd ed.). New York: Routledge.

Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: a case study of causal inference for multilevel observational data. *Journal of the American Statistical Association, 101*, 901–910.

Hurtado, S., Eagan, M. K., Cabrera, N. L., Lin, M. H., Park, J., & Lopez, M. (2008). Training future scientists: predicting first-year minority student participation in health science research. *Research in Higher Education, 49*(2), 126–152.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Thousand Oaks: Sage.

Jones, M. T., Barlow, A. E., & Villarejo, M. (2010). Importance of undergraduate research for minority persistence and achievement in biology. *The Journal of Higher Education, 81*(1), 82–115.

Kilgo, C. A., & Pascarella, E. T. (2016). Does independent research with a faculty member enhance four-year graduation and graduate/professional degree plans? Convergent results with different analytical methods. *Higher Education, 71*(4), 575–592.

Kilgo, C. A., Sheets, J. K. E., & Pascarella, E. T. (2014). The link between high-impact practices and student learning: some longitudinal evidence. *Higher Education, 69*(4), 509–525.

Kim, M. M., & Conrad, C. F. (2006). The impact of historically Black colleges and universities on the academic success of African-American students. *Research in Higher Education, 47*(4), 399–427.

Kim, Y. K., & Sax, L. J. (2009). Student–faculty interaction in research universities: differences by student gender, race, social class, and first-generation status. *Research in Higher Education, 50*(5), 437–459.

Kinzie, J., Gonyea, R., Kuh, G. D., Umbach, P. D., Blaich, C., & Korkmaz, A. (2007). *The relationship between gender and student engagement in college*. Louisville: Paper presented at the annual meeting of the Association for the Study of Higher Education.

Kuh, G. D. (2008). *High-impact educational practices: what they are, who has access to them, and why they matter*. Washington, DC: Association of American Colleges and Universities.

Kuh, G. D., Chen, P. D., Nelson Laird, T. F., & Gonyea, R. M. (2007). Teacher-scholars and student engagement: some insights from FSSE and NSSE. American Council of Learned Societies (ACLS), student learning and faculty research: connecting teaching and scholarship. A Teagle foundation white paper from the ACLS Teagle Foundation Working Group on the Teacher-Scholar. New York: American Council of Learned Societies. http://education.indiana.edu/dotnetforms/Profile.aspx?u=tflaird

Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education, 79*(5), 540–563.

Lopatto, D. (2006). Undergraduate research as a catalyst for liberal learning. *Peer Review, 8*(1), 22–25.

Lopatto, D. (2010). Undergraduate research as a high-impact student experience. *Peer Review, 12*(2), 27.

Mayhew, M. J., Rockenbach, A. N., Bowman, N. A., Seifert, T. A., Wolniak, G. C., Pascarella, E. T., & Terenzini, P. T. (2016). *How college affects students (volume 3): 21st century evidence that higher education works*. San Francisco: Jossey-Bass.

Merkel, C. A. (2003). Undergraduate research at the research universities. *New Directions for Teaching and Learning, 93*, 39–53.

Nagda, B. A., Gregerman, S. R., Jonides, J., von Hippel, W., & Lerner, J. S. (1998). Undergraduate student-faculty research partnerships affect student retention. *Review of Higher Education, 22*(1), 55–72.

Pace, C. R. (1982). *Achievement and the quality of student effort: report prepared for the National Commission on Excellence in Education*. Los Angeles: Higher Education Research Institute, University of California at Los Angeles.

Patrick, A. R., Schneeweiss, S., Brookhart, M. A., Glynn, R. J., Rothman, K. J., Avorn, J., & Stürmer, T. (2011). The implications of propensity score variable selection strategies in pharmacoepidemiology—an empirical illustration. *Pharmacoepidemiology and Drug Safety, 20*(6), 551–559.

Radford, A. W., Berkner, L., Wheeless, S. C., & Shepherd, B. (2010). *Persistence and attainment of 2003-04 beginning postsecondary students: after 6 years (NCES 2011-151)*. Washington, DC: U.S. Department of Education.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Newbury Park: Sage.

Seymour, E., Hunter, A. B., Laursen, S. L., & DeAntoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: first findings from a three-year study. *Science Education, 88*(4), 493–534.

Steiner, P. M., Cook, T. D., Li, W., & Clark, M. H. (2015). Bias reduction in quasi-experiments with little selection theory but many covariates. *Journal of Research on Educational Effectiveness, 8*, 552–576.

Vaughan, A. L., Lalonde, T. L., & Jenkins-Guarnieri, M. A. (2014). Assessing student achievement in large-scale educational programs using hierarchical propensity scores. *Research in Higher Education, 55*, 564–580.

Wang, Q. (2015). Propensity score matching on multilevel data. In W. Pan & H. Bai (Eds.), *Propensity score analysis: fundamentals and developments* (pp. 217–235). New York: Guilford.

Zydney, A. L., Bennett, J. S., Shahid, A., & Bauer, K. W. (2002). Impact of undergraduate research experience in engineering. *Journal of Engineering Education, 91*(2), 151–157.