



Semester Course Load and Student Performance

Nick Huntington-Klein¹ · Andrew Gill¹

Received: 28 September 2019 / Accepted: 6 October 2020 / Published online: 18 October 2020
© Springer Nature B.V. 2020

Abstract

Many college students in the United States take longer than four years to complete their bachelor's degrees. Long time-to-degree can increase higher education costs by billions. Time-to-degree can be reduced if students take more credits each term. While academic momentum theory suggests that additional credits may also improve student performance, and there is a strong positive correlation between course load and student performance, high course load may reduce time investment in each course, giving high course load a negative causal effect on performance. Concern about the negative impact of course load on performance, especially for struggling students, may lead to pushback against policies to reduce time-to-degree by increasing course load. Using longitudinal data from a regional four-year university with a high average time-to-degree, we find no evidence that high course loads have a negative impact on student grades, even for students at the low end of the performance distribution. This result is consistent with a model where students substitute time away from non-education activities when their course loads increase.

Keywords Course load · Time-to-degree · Student performance · Grades

Introduction

Increased time-to-degree from post-secondary institutions in the United States has taken a prominent position along with low completion rates, access, affordability, and mounting student debt as a major public-policy concern in higher education. Among first-time full-time students seeking a bachelor's degree and commencing their studies in 2009, only 39.8% graduated from their first institution attended within 4 years (U.S. Department of Education 2017). The problem is especially severe at public institutions, where the same figure is 34.8% (58.6% at 6 years). The rate at private not-for-profit private institutions is 53.0% (65.6% at six years).

✉ Andrew Gill
agill@fullerton.edu

Nick Huntington-Klein
nhuntington-klein@fullerton.edu

¹ Department of Economics, California State University Fullerton CSU Fullerton, 800 N State College Blvd, Fullerton, CA 92831, USA

These low four and six-year graduation rates are, in part, a reflection of long time-to-degree for successful graduates. The average time enrolled for bachelor's degree completion between July 2014 and June 2015 was 5.2 years in public institutions and 4.8 years in private not-for-profit institutions (Shapiro et al. 2016, Appendix C: Data Tables). Among those receiving bachelor's degrees in public institutions, 29.9% were enrolled for 6 years and 18.2% were enrolled for 7 to 8 years. Of all graduates, approximately 306,000 bachelor's degree recipients were enrolled for 6 years and 186,000 were enrolled for roughly 7.5 years.

The extra time to complete a bachelor's degree can be costly. Comparing the wages of current college enrollees aged 20–25 against current BA holders aged 20–25 from the Current Population Survey, each additional year represents on average \$23,081 in lost earnings, for a rough total of \$29.2 billion in annual foregone wages. In addition, there may be as much as \$12.5 billion in additional tuition outlay, given a the full-time average at public institutions of \$9,970 (College Board 2018) These calculations do not include the expenditures made by the government on education. There can be additional labor market costs if long time to degree is perceived as a negative labor market signal (Witteveen and Attewell 2019; Brugiavini et al. 2020).

As we discuss below, universities could reduce time-to-degree by encouraging students to take more credits (Attewell and Monaghan 2016). Many universities throughout the United States have embarked on such a policy (see, for example, Venit 2017). There are, however, potential unintended negative consequences. Taking additional credits could crowd out study time per course, harming student performance. If, on the other hand, there is no discernable effect of credits attempted on student performance, or if, under the concept of academic momentum (Attewell, Heil, and Reisel 2012), additional credits actually improve performance and persistence by aligning students with their education, advising students to take more credits could be a cost-effective method for improving time-to-degree.

The present day COVID 19 pandemic has altered the academic landscape and created additional challenges to course taking and timely degree completion. Challenges for students include necessary hardware and software requirements for synchronous and asynchronous online instruction and disruptions to family and work life. At the extensive margin, students and their families are making enrollment decisions that will alter their academic trajectories. According to the United States Census Bureau's Household Pulse Survey, a staggering 80.9% of adults in households where at least one adult was planning on taking post-secondary classes this fall have seen their plans cancelled or changed (United States Census Bureau, 2020, <https://www.census.gov/data-tools/demo/hhp/#/?measures=CSR>). At the intensive margin, external pressures related the pandemic including work pressures, housing insecurity, and food insecurity may result in fewer classes attempted per semester. Student engagement may suffer when courses are delivered online. The challenges related to COVID-19 reinforce the importance of the issues we raise in the current study regarding timely degree completion. At the same time, it is not clear that the results in this study generalize to such an unusual situation, and a setting in which a large majority of students are forced to take classes online. University policy will have to address these and other challenges, with student support services and advising likely taking center stage.

In this paper, we investigate the causal effect of course load on grades using a rich set of administrative data from a large four-year public university campus in California. This university has low four-year completion rates, and is currently pursuing a major shift in policy called "Graduation Initiative 2025" to improve those rates (California State University 2019), with higher course loads—taking 15 credits specifically, rather than 12—as a

central part of the policy. Similar initiatives are in place at many universities throughout the United States, including the University of Hawaii (Venit 2017), the University of South Dakota (<https://www.usd.edu/acpc/finish-in-four>), and Oklahoma State University (<https://cas.okstate.edu/advising/finish-in-four>) to name only a few.

Assessing whether course load affects student performance is complicated by the endogeneity of course load. In one empirical strategy, we isolate within-student variation and compare students' grades in semesters when they take 15+ credits against grades in semesters when they take fewer credits. However, our use of student fixed effects is unlikely to be enough to identify a causal effect, and we take four additional approaches to support and refine our causal interpretation. We examine the institutional context and find that a lot of variation in course load is driven by exogenous registration bottlenecks, we control for dynamic academic pressures, we use a time-varying simulated omitted predictor to calculate Rosenbaum (2002)-like bounds for our estimate, and we examine coefficient stability using the methods in Oster (2019) and Cinelli & Hazlett (2020). In addition, to assess whether course load may affect struggling students differently, we examine whether the effect of course load on student performance varies along the GPA distribution using fixed-effects quantile regression methods (Machado & Santos Silva 2019).

Our results provide no evidence that taking 15 credits rather than 12 credits harms student performance. Controlling for student and class-standing fixed effects, we find that taking one additional course per semester leads to a 0.011 standard deviation increase in course grade (on a four-point scale), rather than a negative effect of any size.

Our 0.011 estimate may be biased upwards by omitted time-varying pressures like family, work in the labor market, or health problems. So we do not report as our conclusion that a full course load improves grades. However, under reasonable assumptions we can reject that a full course load has a meaningfully large negative effect on grades, where a "meaningfully large effect" is about half the difference between a B and a B+. We include observed time-varying controls and do not find that our results become negative. We bound the sensitivity of our results to unobserved confounders in three different ways: by simulation, and by following the methods of Oster (2019) and Cinelli & Hazlett (2020). In each case we find that an unobserved confounder could plausibly reduce our positive result to 0 or a small negative value, but that the unobserved confounder would need to be very strong to produce a meaningfully large negative result.

Because our results reject the presence of a meaningful negative effect, they support a conclusion that policies encouraging 15 credits may effectively decrease time-to-degree without harming student performance. A policy encouraging students to take more credits is in contrast to other measures to improve completion time discussed in the literature, such as improving student preparedness. Such policies tend to focus on reducing course failure, where much has been done and the marginal benefit of additional resources is likely to be low. In addition, a policy encouraging students to take more credits has appeal to universities because, relative to measures such as improving student preparedness or reducing student financial need, it can be implemented quickly and at low cost.

Literature

This section reviews some of the relevant literature, focusing on the theoretical explanations for student performance and course load, as well as empirical evidence on those topics.

Factors Affecting Time to Degree

Common explanations for low graduation rates and increased time-to-degree completion typically center on factors such as student preparedness and performance, financial need, and reduced institutional resources (Ishitani 2006; Bettinger and Long 2009; Bound, Lovenheim, and Turner 2010; 2012; Shapiro et al. 2016; Evans et al. 2017; Yue and Fu 2017; Deming and Walters, 2017; Sanabria et al. 2020). Regarding financial need, for example, a portion of this literature investigates the effect of merit-based scholarships on completion rates and time to completion with mixed results (Cornwell, Lee, and Mustard 2005; Dynarski 2008; Scott-Clayton 2011). Students themselves focus on slightly different concerns regarding increased time-to-degree completion. Moore and Tan (2018) found that students see course availability and work and family obligations as primary barriers to timely degree completion.

Another important predictor of degree completion and time-to-degree is “enrollment intensity,” defined as the number of credits taken in a given term (Volkwein and Lorang 1996; Knight 2004; Herzog 2006; Attewell, Heil, and Reisel 2012; Attewell and Monaghan 2016; Shapiro et al. 2016; Witteveen and Attewell 2019). Almost by necessity, a reduction in time-to-degree will require higher enrollment intensity. Completing a typical 120-credit degree program in four years on a semester system requires that a non-remedial student complete 15 units per semester. Yet, it is not uncommon for universities to allow students to maintain full-time status while taking 12 units per semester (Volkwein and Lorang 1996; Knight 2004). Policies allowing 12 units per semester stem in part from a recognition that students face home and work time pressures (Horton 2016), but they can increase time-to-degree.

Advising students to take more credits may be counterproductive if grades suffer and classes have to be repeated, or if many students are unable to handle a full course load and drop out as a result. In what follows, we consider literature relevant to question of students’ college performance with particular attention to GPA. We begin with general theoretical considerations followed by a brief review of empirical evidence for those considerations.

Determinants of GPA: Theoretical Considerations

The workhorse for much of the literature exploring student outcomes, such as GPA, is the so-called education production function (Hanushek 1979, 2020). In this framework, output, which is some measurable student outcome, is related to a set of educational inputs. These inputs traditionally include student characteristics in the form of academic skill, generally measured using high school GPA and standardized test scores (Cohn et al. 2004; Schmitt et al. 2009; Cyrenne and Chan 2012; Danilowicz-Gösele et al. 2017; Caviglia-Harris and Maier 2020), as well as factors such as complex problem-solving skills (Stadler et al. 2018). Traditional inputs also include school characteristics and resources such as teacher skill or school funding. Student course load, which will be discussed more in detail in Sect. 2.4, fits into this theory as related to student effort and time, both of which are inputs to educational outcomes.

Recent work in economics, educational psychology, and higher education have expanded and refined the list of student-based inputs affecting educational attainment and academic success to include the importance of personality characteristics (Schmitt et al. 2009; Heckman and Kautz 2012; Richardson et al. 2012; Kautz et al. 2014; Caviglia-Harris

and Maier 2020; Sweet et al. 2019; and Martínez et al. 2019), peers (Berthelon et al. 2019; Pu et al. 2020) and student health (Larson et al. 2016; and Wallis et al. 2019). Each of these categories of determinants represent a wide literature covering specific cases from practical intelligence (Schmitt et al. 2009), grit (Fosnacht et al. 2019), substance abuse (Wallis et al. 2019), and stress from financial or family difficulties (Larson et al. 2016).

These student-based contributors to GPA combine, in educational production function theory, with school-based contributors. The traditional set of education production inputs on the school side, which includes instructor and instruction quality as well as institutional capacity, has expanded to include student support and engagement-enhancement services, which are expected to improve student performance. Examples of student-engagement services can include first-year seminars (Culver and Bowman 2020), international exchange programs (Brugiavini et al. (2020), and living-learning communities (Caviglia-Harris and Maier 2020). Traditional funding-based inputs are still expected to apply, and can contribute to institutional support for access to technology (Fairlie 2012).

Evidence on the Determinants of GPA

Evidence strongly supports the conclusion that prior academic success and intelligence-based measurements reflect positive inputs to later academic performance. There is a consistent positive association between factors such as high school GPA, SAT, and ACT scores, and high-school percentile rank and college GPA (e.g. Cohn et al. 2004; Schmitt et al. 2009; Cyrenne and Chan. 2012; Danilowicz-Gösele et al. 2017; and Caviglia-Harris and Maier 2020). An interesting feature of the results reported by Cyrenne and Chan is that the positive association between high school GPA and college GPA is enhanced by high school financial expenditures, neighborhood effects and financial resources (Cyrenne and Chan 2012, p. 535). Stadler et al. (2018) found that GPA was positively associated with complex problem solving skills.

Intelligence-based measures are not the only student-based factors correlated with student success. Personality characteristics are also shown to be associated with college GPA. Richardson et al. (2012) find that conscientiousness is the strongest predictor of GPA among standard personality traits. Results on grit are mixed. In a meta-analysis, Credé et al. (2017) find a strong role for perseverance in explaining variation in GPA, but almost no incremental role for grit, after controlling for conscientiousness. Most recently Fosnacht et al. (2019) find that the perseverance subscale of grit was positively related to GPA for college students. Caviglia-Harris and Maier (2020) find that grades are positively associated with conscientiousness in early semesters and with grit up until students' final year. Sweet et al. (2019) and Martínez et al. (2019) find a positive relationship between psychological capital and college GPA.

There has also been attention in the literature to students' health as a determinant of performance. Regarding student health and academic performance, Larson et al. (2016) and Wallis et al. (2019) both find strong relationships between GPA and psychological and physical health. Correlates include mental health history and marijuana use.

Institutions can affect GPA through access to technology and engagement-enhancement programs. Fairlie (2012) provides experimental evidence that access to technology in the form of free computers increases grades and course completion among minority students. Caviglia-Harris and Maier (2020) find that living-learning communities are positively associated with retention, but find no association with a student's GPA. Culver and Bowman (2020) find no effect of first-year seminars on students' grades. The lessons learned

from this literature are that variation in student performance is complex and that understanding this variation requires careful attention to both intelligence-based measurement and students' personality and health. Attention to students' physical and psychological health is also warranted. There is a role for universities to improve performance through access to technology and support services. A focus on student course load may be one determinant of educational production but it is far from the only one.

Course Load and Student Success: Theoretical Considerations

What might explain the relationship between student achievement and unit taking? The literature identifies two avenues by which a heavier course load could affect a student's performance. The first is a basic time allocation problem. The more courses students take, the less time they have to spend on each course. If performance is any increasing function of effort (Stinebrickner and Stinebrickner 2004, 2008; Pu et al. 2020), this implies a negative effect of course load on performance. One consequence of this relationship between effort and time is that other demands on students' time such as employment and extra curricular activities could affect performance via study time. We review the empirical evidence on the potential negative time effects of additional time demands on learning in Sect. 2.5 below.

The second theoretical framework, which implies a positive relationship between course load and student performance, focuses on the level of student involvement. Students who take more classes may take their heavy load as an opportunity to focus more exclusively on school, responding to additional academic responsibilities by substituting away from other activities. "Academic momentum" (Adelman 1999; 2006) suggests that students who complete more credits in their first year at college are more likely to graduate. Attewell, Heil, and Reisel (2012) outline possible mechanisms. They show that students who begin with heavier course loads display a greater level of commitment to their academic goals and studies, and that positive accomplishments early in one's college career engender future successes by promoting individual "self-efficacy and/or academic self-concept" (Attewell, Heil, and Reisel 2012, 28).

Evidence Related to Course Load and Student Success

In this section we review the direct and indirect evidence on how course load affects student success. We begin with the theoretical time allocation mechanism. There is consistent evidence that time investment affects performance. Stinebrickner and Stinebrickner (2004) find a strong positive association between student study time and first-year grades. Stinebrickner and Stinebrickner (2008) and Pu et al. (2020) use exogenous assignment of environmental factors to show that student time and effort, respectively, improve student participation and performance.

The positive effects of study time imply an additional cost to other demands on student time while in college, such as employment. However, the evidence on whether the addition of new responsibilities actually reduces study time is mixed. Employment has weak crowding-out results in both high school (Kalenkoski and Pabilonia, 2009) and college (Babcock and Marks, 2011).

Weak effects of employment on actual study time are accompanied by inconsistent evidence of the effects of employment on performance. Results among high school students are inconsistent across settings, students, and specifications (Ruhm 1997; Rothstein 2007).

At the college level, the evidence is again mixed. Ehrenberg and Sherman (1987) find no detrimental effects of work on grades, but do find that persistence and time to completion are adversely affected. Stinebrickner and Stinebrickner (2003) show that increasing hours of work by one hour per week reduces semester GPA by 0.162, identified by variation in work hours in assigned required jobs at a small liberal arts college. Darolia (2014), in contrast, finds that working reduces credits completed but does not harm grades. Sports and other extracurricular activities are another possible avenue where time commitments could crowd out study time and reduce grades. However, athletic or college-gym participation does not appear to have strong direct effects on grades, and in some cases may increase them by increasing campus visits (Emerson, Brooks, and McKenzie 2009; Fricke, Lechner, and Steinmayr 2018). Thus, while the available evidence supports a positive relationship between study time and student performance, evidence for both high school and college students provide limited evidence that work and other time demands harm academic performance.

Inconsistent results may be due to whether additional time demands are strong enough to actually reduce study time, or whether they may drive students to visit campus more often, as in Fricke, Lechner, & Steinmayr (2018). Students who take on additional responsibilities, at least at low levels, may be able to substitute from other tasks besides study time.

The available evidence on academic momentum, typically identified by matching on observables, broadly supports a positive association between the completion of first-semester or first-year credits and student success. More intense early enrollment is associated with improved degree completion at two- and four-year colleges, especially among minority students (Attewell, Heil, and Reisel, 2012; Attewell and Monaghan, 2016). These effects also apply to grades (Brugiavini et al. 2020), and effects may be stronger if intensity is maintained (Belfield et al. 2016). Volkwein and Lorang (1996), Knight (2004), Belfield et al. (2016), and Venit (2017) all find direct or indirect evidence that increase course load reduces completion time, although Cornwell, Lee, and Mustard (2005) find only mixed evidence.

Several successful large-scale policy implementations, intended to improve completion by a number of simultaneous interventions, include increased course loads among other interventions. Scott-Clayton (2011) finds that a merit-based scholarship program, which implemented a course load minimum, improved both earned credits by the end of freshman year and BA completion rates, although the effects on completion faded out (Scott-Clayton and Zafar 2016). Denning (2018) finds more general evidence that financial aid increases credits taken and reduces time to degree, while also increasing graduation rates. Scrivener et al. (2015) report on a randomized controlled trial including a requirement to attend school full time, tuition waivers to cover residual differences in financial aid and tuition and fees, and encouragement to “take developmental classes early and to graduate within three years.” (Scrivener et al. 2015, iii). The authors report very large treatment effects, on the order of an 18 percentage-point (82%) increase in completion rates, as well as improvements in completion times. Taken together, this literature supports a conclusion that retention and performance improves when students are more focused on school by taking more credits.

The bulk of this literature informing the question of whether increased course load harms performance is indirect, looking at other demands on time, on the impact of course load on academic focus, or the impact of policies that increase course load but also change other things such that the impact of course load cannot be separated from other effects. There is a small amount of direct evidence on the effect of course load on retention and grades, although it is generally non-causal. Szafran (2001) and Jackson et al. (2003) report a associations between first-semester course load, retention, and grades. Venit (2017),

reporting on the University of Hawaii's "15 to Finish" program, found no evidence that student performance was harmed when taking a 15 credit course load, and found that taking at least 15 credits per term in the first year, as opposed to fewer, was associated with persistence and GPA, even for the lowest performing high-school students.

The evidence on increased course-load appears to tilt in the favor of the academic momentum theory rather than any sort of time allocation explanation in which time use of non-academic tasks is inelastic. However, much of this literature is based on raw correlation or selection on observables, with estimates from matching as the forefront of the causal side of the field when studying course load alone rather than as a part of an RCT package. Further, the evidence on student grades is scant. We address this part of the literature by using detailed administrative data that allows for a fixed-effects design and the observation of student grades.

Data

We use administrative data from a major four-year university, provided by the office of Institutional Research & Analytical Studies (IRAS). The university in question is a regional, moderately selective four-year university and is one of the 23 California State University (CSU) campuses. It serves a wide range of students, including a high proportion of Hispanic students. It is important to note that the CSU system and our own university enacted a policy to recommend higher course load, but that this occurred after our sample window.

Data are at the student-course level and include information on all courses attempted, including those that the student failed or withdrew from, and grades received for two incoming cohorts of freshmen students, 3,874 students beginning Fall 2010 and 4,141 beginning Fall 2011.¹ To ensure consistency in estimates, we omitted the small number of students who did not have full demographic information. Our particular interest is in analyzing the effect of course load on student performance among students enrolled full-time at the margin between taking a full course load or not, so we drop part-time terms. There are two reasons for this restriction. First, we do not want to pick up potential nonlinearities in the effect of course load below 12 credits. Second, our focus on student performance for those students moving from 12 and 15 credits speaks directly to policy prescriptions advocated our own university and by other institutions (e.g. Venit 2017) to reduce time to completion. One potential downside of dropping part-time terms is that some of these students may have co-enrolled for additional credits at a community college. Our administrative data do not contain information on these additional credits taken.

Additional restrictions include dropping summer terms, even though summer classes may be one way of graduating on time without taking a full course load, because students generally take only one or two classes at a time in the summer, at triple speed, and so the impact of an additional class would not be comparable. We also drop students who only took classes with extremely low variation in grades (described below). These restrictions result in very few students being dropped from the sample.² Our final samples sizes for

¹ Our Institutional Review Board (IRB) approved our research protocol and methods for handling and storing these administrative data.

² Few students were dropped entirely, but omitting student-terms with fewer than 12 credits led to 6,810 student-terms being dropped. Another 1,650 student-terms were dropped due to a small number of missing

Table 1 Descriptive statistics

	Sample University		National ^a	
	Mean	(SD)	Mean	(SD)
Student-level data:				
Female	0.577		0.562	
Received Financial Aid	0.814		0.740	
Admissions Index	3556.971			
High school GPA	3.324		3.260	(0.451)
Modal term is not full course load	0.533			
Modal term is full course load	0.467			
Has any course load variation ^b	0.808			
White	0.254		0.570	
Black	0.026		0.140	
Hispanic	0.405		0.127	
Asian/Pacific Islander	0.225		0.045	
Other/mixed	0.090		0.092	
Nonresident alien			0.026	
Other levels:				
Grade in class	2.800	(0.778)		
Grade standardized within class	– .011	(0.732)		
Classes attempted in term	4.63	(0.560)		
Non-modal number of units taken this term ^b	0.261			

^a National summary statistics except for GPA are from 2011 IPEDS data for 4-year public or not-for profit private degree granting universities in the United States, offering BA, MA, or PhD as the highest degree. National GPA is from NLSY97 transcript data for individuals aged 18–22 enrolled in a four-year college on September 2002

^b Students with only one 12+ credit term dropped from this calculation

analysis are 3,852 and 4,102 students beginning in Fall 2010 and 2011, respectively. All remaining calculations use this subsample.

We observe student course-taking and grades through the end of the Spring 2017 semester. We observe courses and grades for 18,337 student-terms while the students were freshmen, 13,835 while sophomores, 12,813 while juniors, and 14,199 while seniors. The relationship of sample size to class standing is due to dropout along with students who spend multiple years as seniors. There were 34.4 classes attempted, on average, across all students. 64% of the students from the incoming cohorts had graduated within 6–7 years (by Spring 2017). 26.9% of graduates graduated within four years (by Spring 2014–15).

In addition to course-taking, we observe student background characteristics, including self-reported race/ethnicity and gender, student standing (freshman, sophomore, junior, or senior), financial aid receipt, high school GPA, declared major, and an admissions index based on a combination of high-school GPA and ACT/SAT scores. We also observe

Footnote 2 (continued)

admissions indices, or student-terms in which the only classes taken were non-graded or had grade variance beneath the 5th percentile.

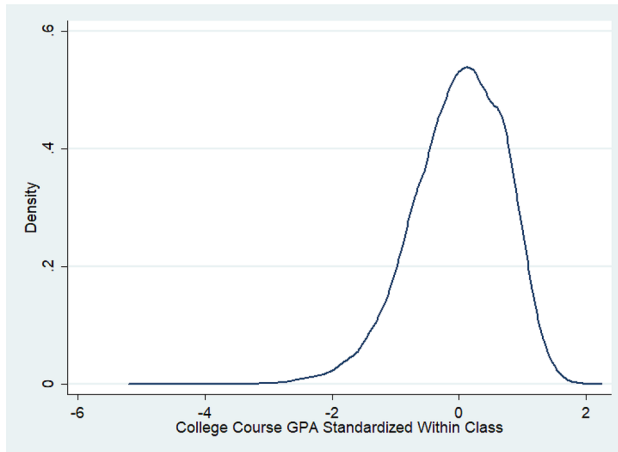


Fig. 1 Distribution of the GPA Standardized Within-Class

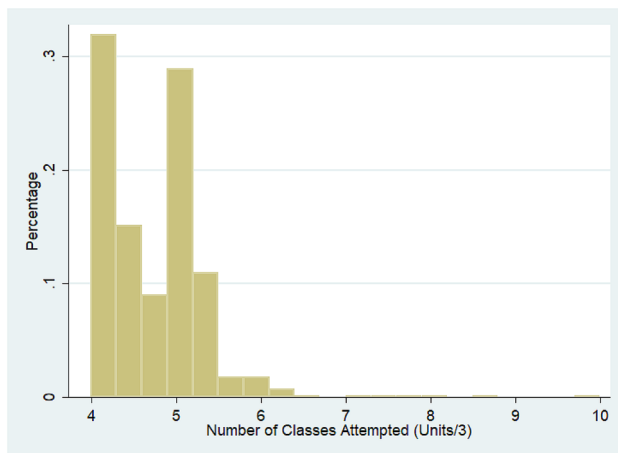


Fig. 2 Distribution of the Number of Classes Attempted per Term

aggregate measures of background and performance by students in other cohorts who were in the same classroom as students in our sample.

Table 1 provides descriptive statistics. A large fraction of the students in the incoming cohorts analyzed report that they are Hispanic (0.405) and a large fraction report receiving some financial aid (0.814). These figures are higher than national averages of Hispanic students (0.127) and financial aid receipt (0.740), as shown in the third column.

Average college course grades are 2.800, with a standard deviation of 0.778. To account for variation in grades driven by between-course differences, as described in the next section, grades are standardized within class section. Since the goal is to explain within-class variation in grades, and to avoid inflated standardized-GPA values, we drop courses in which the within-class standard deviation of grades is below the 5th percentile (a standard

deviation of 0.173 grade points). Figure 1 shows the distribution of within-class standardized GPA.

On average, students attempted 4.63 classes per semester, where taking 5 classes is a full course load and 4 is the minimum necessary to be considered a full-time student. 46.7% of students take less than a full course load in a modal term. Figure 2 shows the full range of course load, without collapsing the data to a full/not-full binary, and shows that across all students and semesters, the modal number of classes attempted was exactly 4, with 5 close behind.

For analysis, we consider a term to be a full course load if the student attempts 5 or more courses (15 or more credits), a partial course load if they attempt fewer than 5 courses (12–14 credits), and since we are interested in the margin between partial and full course load for full-time students, the term is dropped if fewer than 12 credits are taken. In the remaining sample, 44.0% of terms are with full course loads. Course load is determined on the basis of attempted classes, and so students who later drop or fail a course are still counted as taking it.

While the decision to take a full course load has a basis in student-level characteristics (as will be examined later), there is significant within-student variation in the number of courses taken per term, necessary for a fixed effects identification strategy. Excluding students who dropped out after one semester, students take a course load other than their modal course load (i.e. a student who usually takes 15+ credits taking 12–14 instead) in 26.1% of semesters. 80.8% of all students have at least some within-student variation in course load. In the context of the university, there is qualitative evidence suggesting that a fair amount of below-full-load taking behavior occurs because students try to take over-enrolled classes, do not manage to register from the waiting list, and do not replace the class with another one (Moore and Tan 2018). To the extent that this is the driver of within-student variation, concerns about selection bias in estimators using within variation are minimized.

Identification

A clear source of endogeneity in estimating the effect of course load on student performance is ability bias. Higher student ability should both lower the effort cost of taking more classes as well as increase the expected grade performance, leading to a positive non-causal observed relationship between classes taken and grades. In this paper we use within-student variation to account for ability bias.

Putting ability bias aside, we take into account the theoretical explanations covered in Section II. The academic momentum theory implies a positive effect of course load on student performance that, if it exists, should be part of the effect identified. If there are “identity” effects to taking many classes, or if additional classes increase time spent on campus, increasing study time per class, then students will increase study time when they take heavy course loads, driving a positive effect of course load on grades. Time allocation theory also implies a causal effect: if students substitute away from study time on other courses to some degree when they add a new course, course load should reduce performance. However, in addition to its implied causal effect, the time allocation theory suggests several possible biases in the results, which we outline below.

If a student faces a negative shock requiring them to work in the labor market or aid their family more, they may be likely to reduce both the number of classes taken and the effort spent on each class at the same time, driving a positive correlation between classes taken and grades. Similarly, if course load is chosen before effort, then a consumption shock occurring between those two choices may require the student to lower effort more sharply in each class if enrolled in more classes, driving a negative bias.

The bounded nature of course loads can also drive a positive bias. If a student is incentivized to spend less time on school in a given period, but is already taking few classes, they can only reduce effort rather than taking fewer classes. Alternately, if students are planning to take four classes but one is overbooked (Moore and Tan 2018), they must replace the class to remain adequately enrolled. But for a student taking five classes and failing to enroll in one, replacing the overbooked class is optional, and more motivated students may be more likely to do so.

Finally, there is the endogenous feature of the difficulty of courses. Students may choose the number of courses to take on the basis of how difficult they expect their course mix to be (Volkwein and Lorang 1996; Cornwell, Lee, and Mustard 2005). If students take more courses when their courses are easier, the impact of course load will be positively biased.

Identifying the effect of the course load on performance should then focus on within-student variation to account for unmeasured ability. Further, it must take into account time-varying external factors likely to drive the time allocation decision, including the difficulty of the courses being taken in a given term, which is also endogenous.

We describe the relationship between performance GP_{ijt} , measured as the grade points earned in the class by student i taking the mix of classes j in a given term t , and the number of credits taken that term (divided by three to give the number of classes taken). Performance is dictated by individual unobserved components—like ability— β_{0i} , class-specific components that determine grades—like the generosity of the grading scale— β_{1j} , a direct effect of the number of classes taken $Class_{it}$, and individual time-varying determinants that are observed such as class standing or previous grades earned (X_{it}) or unobserved such as family events or consumption shocks (ε_{it}):

$$GP_{ijt} = \beta_{0i} + \beta_{1j} + \beta_2 Class_{it} + \beta_3 X_{it} + \varepsilon_{it}. \quad (1)$$

At the university studied, there is not an enforced grade curve, and so β_{1j} varies across classes. We move β_{1j} to the left side of Eq. (1), using average grades earned by students in the course as a proxy, and replacing the left-hand side with grades standardized within course section:

$$GP_{ijt}^{STD} = \beta_{0i} + \beta_2 Class_{it} + \beta_3 X_{it} + \varepsilon_{it} \quad (2)$$

The number of classes chosen is endogenous, and is determined by some individual fixed factor γ_{0i} , features of the course mix the student is facing (also endogenous) γ_{1j} , and individual time-varying observed (X_{it}) or unobserved components (v_{it}).

$$Class_{it} = \gamma_{0i} + \gamma_{1j} + \gamma_2 X_{it} + v_{it} \quad (3)$$

We use a fixed-effects estimator that accounts for the influence of both β_{0i} and γ_{0i} . Still, since $Class_{it}$ and γ_{1j} are chosen jointly, their influence cannot be separately identified. The estimated coefficient $\hat{\beta}_2$ should be interpreted as a mix of the effects of classes and of course mix. However, this reduced-form effect is of interest, since any impact of a policy that increases course load on how students choose which courses to take should be a part of

Table 2 Classes attempted and class performance

	Dependent variable: standardized class grade				
	(1)	(2)	(3)	(4) Graduates Only	(5) Non-Grads Only
Full course load	0.124*** (0.008)	0.089*** (0.007)	0.011** (0.005)	0.003 (0.006)	0.017 (0.014)
Female		0.011 (0.011)			
Black		-0.204*** (0.036)			
Hispanic		-0.167*** (0.014)			
Asian/PI		-0.035** (0.016)			
Other/mixed		-0.080*** (0.022)			
Admsn Index (Standardized)		0.020*** (0.007)			
Financial aid		-0.026* (0.015)			
High school GPA		0.533*** (0.017)			
Student standing fixed effects	Y	Y	Y	Y	Y
Student fixed effects	N	N	Y	Y	Y
N observations	59,184	59,184	59,184	44,651	14,533
N students			7,954	5,109	2,845
R ²	0.012	0.103	0.009	0.004	0.041

Student standing is freshmen/sophomore/junior/senior. Standard errors clustered within student

the effect of interest. Some analyses will attempt to disentangle the mechanisms at play by controlling for different features of the course mix.

There is not an available instrument to directly account for the correlation between ϵ_{it} and v_{it} , which is likely nonzero because pressures from employment hours are likely to be in both ϵ_{it} and v_{it} . Basic estimates will be biased by this correlation. We estimate partial identification bounds on β_2 to determine the strength of the relationship necessary to change the substantive results of the paper. Additionally, we emphasize that, in the context of the university being studied, a significant driver of variation in course load is the presence of registration bottlenecks (Moore and Tan 2018), which do not present the same problem for identification as do time-varying pressures in the student’s life.

Results

This section gives our empirical results. We first report standard fixed effects estimates investigating the effect of course load on student performance, and then follow with examination of potential observed and unobserved sources of bias in these basic findings.

Fixed Effects Results

Table 2 shows our fixed effects estimates explaining student performance. The dependent variable for all regressions in Table 2 is a student's grade in each class, standardized within the particular class section. Standardization occurs within the section, rather than for the same class across all sections. All standard errors are clustered at the student level. In Column 1 we show pooled OLS results giving the relationship between attempting a full course load in the term and standardized grade, controlling only for a student's class standing (freshman / sophomore / junior / senior). In Column 2 we add student-level characteristics. The within estimates regressing standardized grade on taking a full course load appear in Columns 3–5 of the table, where the Column 3 results are for all students and Columns 4 and 5 results are limited to students who successfully graduated, or did not graduate, respectively.

As shown in Column 1, taking a full course load is associated with a 0.124 standard deviation increase in grade. The association between full course load and grade is reduced by roughly 28% when we add student-level characteristics in Column 2. Controlling for high school GPA and admissions index, all race/ethnic groups show lower standardized grades than whites (the reference category), and a one point increase in high school GPA is associated with a 0.533 standard deviation increase in college course grade.

Column 3 shows a within-student estimate of the effect of classes attempted on grade. The effect is positive, but is so small (0.011) that it is of little practical significance. Importantly, we can reject to a reasonable degree of precision that the within-student relationship between a full course load and GPA is negative. Additionally, Column 3 gives some insight to the likely correlation between fixed unobserved student heterogeneity and classes attempted. The reduction in the estimate from Column 1 to Column 2 to Column 3 implies that it is higher-performing students who do tend to choose higher course loads, and that many of the relevant high-ability characteristics are not measured in our data.³

Columns 4 and 5 of Table 2 show that the effect of a full course load is not meaningfully different for graduates or non-graduates. The effect is not large or significant for either subgroup, but is slightly more positive for non-graduates, who are less likely to choose higher course loads in the first place.

Examining variation in the effect in other ways besides eventual graduation success, we allow for the possibility that the effect of course load on grades differs by key student demographic characteristics. In six separate analyses we interact the full-course-load variable with race/ethnicity (white/black/Asian/Hispanic/other or mixed), gender, high-school GPA (continuous), parental education (no college, some college, graduate), parental income (seven bins), and lagged cumulative GPA (continuous), respectively. In results

³ Results are almost entirely unchanged if observations are weighted by the inverse of the standard deviation of treatment within group to account for the fixed effects estimator overweighting individuals with high treatment variance, as in Gibbons, Serrato, & Urbancic (2018).

Table 3 Quantile regression of the effect of full course load

Percentile:	Dependent Variable: Standardized Class Grade				
	(1) 10 th	(2) 25 th	(3) 50 th	(4) 75 th	(5) 90 th
Full Course Load	0.044*** (0.010)	0.029*** (0.007)	0.009* (0.005)	– 0.008 (0.006)	– 0.020** (0.008)
Student Standing Fixed Effects	Y	Y	Y	Y	Y
Student Fixed Effects	Y	Y	Y	Y	Y
N Observations	59,184	59,184	59,184	44,651	14,533
N Students	7,954	7,954	7,954	7,954	7,954

Student standing is freshmen/sophomore/junior/senior. Standard errors clustered within student

available upon request, we find that none of the coefficients for these interactions are significant (the lowest p -value for a joint test of a set of interactions is 0.348). Among categorical interactions, there is only one case—the second-lowest parental income bin—where the point estimate for that group is negative (although it is insignificant).

Finally, we examine whether the effect varies along the GPA distribution. It is reasonable to expect that high-ability students might have a null or positive effect of increased course load, but that students already having difficulties might not be able to handle the additional work. In Table 3 we use the method described in Machado & Santos Silva (2019) to perform fixed-effects quantile regression. Surprisingly, we find that the effect is most positive at the low end of GPA, and actually turns negative and significant for the 90th percentile. The potential concern that strong students drive the nonnegative result, and that full course loads would still be too much for low-GPA students to handle, is not supported.

Observed Sources of Bias

Student ability is one determinant of course load, which can be addressed using fixed effects, but there may be other important factors. Table 4 demonstrates the predictors of course load, which may help to determine the extent of the known bias in Table 2 and how it might be reduced. All standard errors are clustered at the student level. As shown in Column 1, students taking more classes tend to have higher high school GPAs. Women are more likely to take full course loads. White students are most likely to take a full course load, with black students coming close behind. Hispanic students, about 40% of the sample, are 13.4 percentage points less likely to take a full course load than whites.

The other columns in Table 4 show results giving the influence of time-varying student characteristics on the number of classes taken, and include student fixed effects. As shown in Columns 2 and 3, students tend to take more courses when their courses are easier. However, this is based not on the raw grades given out in class (Average Grade in Course), but rather the grades given out relative to what might be expected based on the demographic characteristics and cumulative GPA of students taking the class (Population-Adjusted Average Grade). Roughly, Population-Adjusted Average Grade can be thought of as “easy for the people who choose to take it.” An average course grade one GPA unit higher than

Table 4 Predicting a full course load

	Dependent variable: full course load				
	(1)	(2)	(3)	(4)	(5)
Female	0.030*** (0.007)				
Black	– 0.020 (0.023)				
Hispanic	– 0.134*** (0.008)				
Asian/PI	– 0.062*** (0.009)				
Other/mixed	– 0.046*** (0.012)				
Admsn Index (Standardized)	0.008** (0.004)				
Financial aid	– 0.017** (0.008)				
High school GPA	0.043*** (0.010)				
Average grade in course		0.005 (0.010)			
Population-adjusted Average grade GPA last term			0.183*** (0.023)		
				0.068*** (0.005)	0.067*** (0.005)
Cumul. GPA Last term					0.104*** (0.012)
Student standing fixed effects	Y	Y	Y	Y	Y
Student fixed effects	N	Y	Y	Y	Y
N observations	59,184	59,184	59,184	44,027	36,594
N students		7,954	7,954	7,601	6,933
R ²	0.023	0.007	0.009	0.014	0.015

Population-adjusted Average Grade is the average course grade earned in that course, adjusting for demographics and cumulative GPAs of students who take the course. “Student standing” is freshmen / sophomore / junior / senior. Standard errors clustered within student. Sample size drops for columns 3–5 to account for lags

could be expected given who takes the course is associated with the student being 18.3 percentage points more likely to take a full course load. Columns 4 and 5, which omit the student’s first term (and first two terms, respectively) examine the student’s recent performance in college. Columns 4 and 5 show that a student who has recently been doing better than usual is likely to take more courses, which could be interpreted as confidence, or learning one’s limits. In all regressions in Table 4, R² values are low, indicating that these measured confounders are not especially strong.

Table 5 Robustness of the classes attempted and class performance link

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: standardized class grade				
			First term excluded	Terms 1 & 2 excl	Terms 1 & 2 excl
Full course	0.011** (0.005)	0.013** (0.005)	0.012** (0.006)	0.018*** (0.007)	0.013* (0.007)
Load	0.041*** (0.014)				-0.001 (0.016)
Average grade in Course		-0.241*** (0.030)			
Pop-adjusted					
Average grade			0.115*** (0.006)	0.079*** (0.007)	0.073*** (0.007)
GPA last term				-0.227*** (0.016)	-0.223*** (0.016)
Cumul. GPA					
Last term				Y	Y
Student standing fixed effects	Y	Y	Y	Y	Y
Student fixed effects	Y	Y	Y	Y	Y
Declared major fixed effects	N	N	N	N	Y
N observations	59,184	59,184	44,027	36,594	36,594
N students	0.010	0.011	0.022	0.021	0.032
R ²	7,954	7,954	7,601	6,933	6,933

Population-adjusted Average Grade is the average course grade earned in that course, adjusting for demographics and cumulative GPAs of students who take the course. “Student standing” is freshmen / sophomore / junior / senior. Standard errors clustered within student. Sample size drops for columns 3–5 to account for lags

Despite low R^2 values in Table 4, Table 5 checks whether any of these predictors can explain the positive relationship between full course loads and grades. All standard errors are clustered at the student level. Columns 1 and 2 display the results when we add average course grades. While the left-hand side variable is already adjusted for average grades in each course, it is still possible that multiple easy classes may allow a student to overperform in all of them at once. However, neither addition eliminates the positive relationship. The effect also is not eliminated by controlling for recent or prior performance in Column 3, indicating that the positive relationship is not a result of students, for example, taking more classes as they discover which fields they are good at. Finally, in Column 5, we include all of the previous controls (using the non-population-adjusted Average Grade, although it does not make much difference) and also include fixed effects for declared major to account for differing institutional standards for how many courses students should take, especially in STEM. The effect remains positive.

Columns 3 and 4 are especially important. As will be discussed in the next section, a potential source of omitted variable bias is time-varying pressures from work or family, which may reduce course load and, at the same time, harm performance in class. It is reasonable to expect that these time-varying pressures may be serially correlated and would also affect grades. If they are, then lagged GPA acts as a proxy for these time-varying pressures, and so Columns 3 and 4 act as a partial test for this source of omitted variable bias, and fail to change the result.

The relationship between courses taken and grades remains positive and statistically significant. While the effect is small, the qualitative result of importance is that the relationship is not negative, even accounting for time-constant skills with fixed effects, the tendency to take many easy courses at once, the tendency to take more courses when one is doing particularly well, or institutional differences between majors.

Unobserved Work Requirement as a Source of Bias

There is an important confounder missing from our analysis in Sect. 5.2—consumption shocks or time-varying work pressures, or similar time-varying pressures from family or other sources. At the setting studied, significant portions of students work either part or full time while attending classes. Theoretically, we would expect that changes in course load driven by factors that increase work hours would reduce grades and courses taken at the same time, leading to a positive bias. Under the assumption that these pressures are serially correlated, lagged GPA is a proxy for these time-varying pressures, and in Table 5 a control for lagged GPA does not change the result. Still, controlling for lagged GPA is unlikely to eliminate all bias.

Since we do not observe work hours, we perform a simulated sensitivity test, inspired by Rosenbaum (2002) bounds. We randomly generate a binary unobserved confounder Z_{it} , which could be considered to represent something like a consumption shock or outside work demands, using the formula

$$Z_{it} = I(\delta_1 GPA_{it}^{STD} + \delta_2 (I(Class_{it} \geq 5) - .5) \geq \mu_{it}),$$

where $\mu_{it} \sim N(0, 1)$. By construction, Z_{it} is positively correlated with both GPA_{it}^{STD} and with $Class_{it}$, unless δ_1 or $\delta_2 = 0$. We allow δ_1 and δ_2 to each take any value $\{0, 0.025, 0.05, \dots, 1\}$, which generates correlations between Z_{it} and GPA_{it}^{STD} of between 0 and 0.323, and correlations between Z_{it} and $Class_{it}$ of between 0 and 0.46. We then add Z_{it} as a predictor to the model, and examine how the coefficient on $Class_{it}$ changes as a result. We repeat

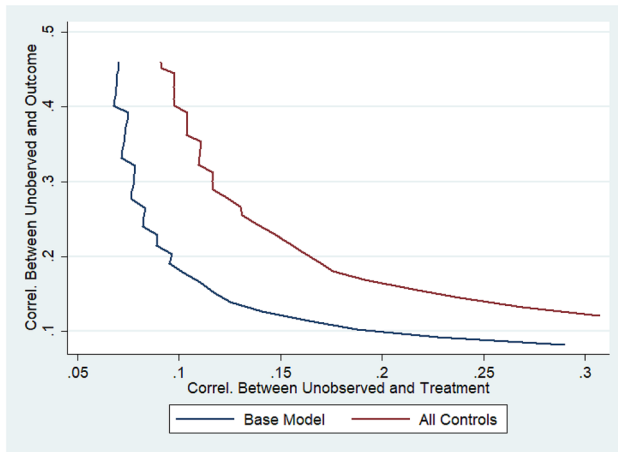


Fig. 3 Minimal Correlation with Omitted Predictor Required to Generate Negative Significant Result. Simulation described in Sect. 5.3. Lines indicate the correlation between a simulated unobserved variable and treatment/outcome that generates a negative result in 95% or more of simulations. Slight positive slopes are due to indirect manipulation of correlation and a discrete search space

this process 200 times for each combination of δ_1 and δ_2 , and for each of two models: the baseline model (Table 2 Column 3), and including all controls (average grade, prior performance, and current declared major). The goal is to determine the strength of the relationship between Z_{it} and the observed variables necessary to generate a significant negative coefficient on $Class_{it}$.

Figure 3 shows the minimum correlations necessary to generate a statistically significant negative coefficient on $Class_{it}$ at the 95% level (190 or more out of 200 simulations are negative). In other words, for this analysis to be heavily biased enough by an omitted predictor such that we should be reporting a negative effect of number-of-classes on grade performance, that omitted predictor would need to have a correlation strength on the drawn boundary or to the top-right of it. In the base model without additional controls, such an omitted predictor would need to have a correlation of about 0.1 or better with GPA_{it}^{STD} and a correlation of about 0.2 or better with $Class_{it}$, although one stronger correlation can trade off for another.

Correlations in the range of .1 – .2 between time-varying work or family pressures and course load or GPA are not particularly high, indicating that an omitted predictor needs to have only moderate strength to mask a statistically significant negative result. However, even in these statistically significant cases, the negative relationship between $Class_{it}$ and GPA_{it}^{STD} is not large. Considering only statistically significant negative coefficients in the simulation, the average coefficient on $Class_{it}$ was -0.032 , and ranged from -0.084 to -0.009 .

To determine the strength of omitted predictor that would be necessary to produce a *meaningfully large* negative effect, we increase δ_1 and δ_2 until we find meaningfully large negative effects. We would consider a 0.2 drop in within-class standard deviations of GPA to be a moderate effect size. In absolute GPA terms this translates to about 0.154 GPA points, about half of the difference between a B and B+. We achieve this moderate negative effect by searching values of δ_1 and δ_2 in $\{2.2, 2.25, \dots, 2.7\}$ and find moderately large effects of -0.2 with rather large within-student correlations of at least 0.399 with course

load and 0.531 with GPA, with a small amount of potential tradeoff between the two. With controls included, correlations need to be 0.430 with course load and 0.531 with GPA. We would consider these to be fairly strong within-student correlations, possibly implausibly high.

In addition to the simulation of Rosenbaum-like bounds, we also examine coefficient stability using the method in Oster (2019). We make the assumption that observable (using the full model above, average grade, prior performance, and current declared major) and unobservable characteristics have identical levels of relative influence on selection into a full course load on the basis of observables and impact on GPA ($\delta = 1$ in the terms Oster uses, see Cinelli & Hazlett (2020) for detail on the interpretation of δ). Given this assumption, we check every R^2 value in steps of 0.05 to calculate the minimum R^2 of a model including both observables and unobservables that would lead to either a negative effect size (< 0) or a meaningfully large negative effect size (< -0.2). If the model including both observable and unobservable controls has a within-student R^2 of 0.1 or greater, then the estimated effect is negative. However, to produce an effect size that is meaningfully large and negative, the full model would need to have a within-student R^2 of 0.6 or greater, which is high.

We then use the coefficient stability methods of Cinelli & Hazlett (2020), which, like Oster (2019), constructs bounds by comparing the influence of observed and unobserved controls. With all controls included, unobserved controls would need to have variation orthogonal to the included controls that explains at least 17.4% of both the residual variation in the treatment and residual variation in the control to produce a point estimate of -0.2 . There is some possibility for tradeoff between explaining residual variation in the treatment and the control. Further, Cinelli & Hazlett (2020) provide a means of bounding the adjusted confidence interval of an unobservable-including estimation using observed covariates. We check all included covariates and find that the adjusted confidence interval never includes -0.2 , and at its lowest is -0.005 .

We conclude that, even though our main results likely contain positive omitted variable bias, and we would not be surprised to find a small negative effect of course load with the proper controls, our partial identification analysis shows that adjusting for this bias is unlikely to lead to a meaningfully large negative causal relationship between $Class_{it}$ and GPA_{it} , which is consistent with our broader conclusion that the effect of course load is meaningfully near zero. For a large negative effect to emerge with proper controls, time-varying outside pressures would need to be extremely strong determinants of course load, which contradicts our earlier discussion that in this context, much of the course load decision has to do with course bottlenecks outside the control of the student.

Supplementary Results

In Table 6 we show some supplementary results related to the effects of taking more classes each term. In Columns 1 and 2 we predict persistence to the next term. These results address the possibility that additional classes, even if they do not weaken performance, may lead to burnout so that students are less likely to return. Consistent with much of the literature in Sect. 2, we find no such relationship.

Columns 3 and 4, which are performed on a one-observation-per-student basis, examine the relationship between taking more classes per term and the rate of graduation as well as the time to graduation. The variable “Proportion Full Course Loads” is an average over all enrolled terms. Without any within-student variation, the causal identification for

Table 6 The relationship between classes attempted, persistence, and completion

	Persist to next term (Logit Marginal Effect)		Graduate (Logit Marginal Effect)	Time to grad. (grads only)
	(1)	(2)	(3)	(4)
Full course	0.012***	0.020***		
Load	(0.002)	(0.006)		
Proportion full			0.310***	– 1.099***
Course loads			(0.016)	(0.032)
Female	– 0.001		0.042***	– 0.185***
	(0.002)		(0.010)	(0.017)
Black	0.001		– 0.048	0.078
	(0.005)		(0.033)	(0.057)
Hispanic	0.010***		0.018	0.064***
	(0.002)		(0.013)	(0.022)
Asian/PI	0.013***		0.049***	0.103***
	(0.003)		(0.014)	(0.024)
Other/mixed	0.001		– 0.002	0.036
	(0.004)		(0.019)	(0.032)
Admission Index (Standardized)	0.001		– 0.002	– 0.007
	(0.001)		(0.006)	(0.009)
Financial aid	– 0.001		– 0.010	– 0.002
	(0.002)		(0.013)	(0.022)
High school GPA	– 0.003		0.104***	– 0.146***
	(0.003)		(0.016)	(0.027)
GPA this term	0.039***	0.043***		
	(0.001)	(0.005)		
Cumulative GPA	0.029***	0.190***		
	(0.002)	(0.021)		
GPA first term			0.153***	– 0.144***
			(0.006)	(0.014)
STEM major ever	0.003		– 0.057***	0.216***
	(0.002)		(0.013)	(0.025)
Student standing fixed effects	Y	Y	N	N
Student fixed effects	N	Y	N	N
N observations	51,261	51,261	7,954	5,109
N students		7,952		

Column 3 and 4 estimates are noncausal. “Student standing” is freshmen/sophomore/junior/senior

these estimates is very weak, and requires that the list of controls is sufficient. So we consider these non-causal. There is a very strong relationship between taking more classes and graduating; students who take one more class per term graduate more than 30 percentage points more often. There is also a negative relationship between classes per term and time-to-degree. Students taking an additional class each term take on average 1.1 fewer years to graduate, more than the full year decrease one might expect mechanically.

Finally, there are one-unit classes and a small number of non-3-unit classes in the data. We ran all results again using a more direct count of classes (excluding one-credit classes) rather than the number of units divided by 3. In results available upon request, we find that the findings are very similar, and while point estimates differ slightly (and some of the positive-and-significant-but-very-small results that we emphasized as being upper bounds became positive-and-insignificant) none of the substantive conclusions about the effect of class taking changed. Two results did change: (1) the effect of class-taking for the top decile of students changes from negative and significant to negative and insignificant, and (2) results were different for Table 4, in regards to predicting which students take full course loads. In particular, racial and gender effects become less prominent in Column 1, and the influence of average class grades is now negative (Columns 2 and 3).

Discussion

The negatives of long time-to-degree are clear: earnings penalties for some (Witteveen and Attewell 2019), poorer overall performance, perhaps due to skill atrophy (Brugiavini et al. 2020), cohort crowding (Kurlaender et al. 2014), and additional financial cost in terms of tuition outlays and delayed entry into the labor market. One way to improve time to degree would be for institutions to support efforts to increase credits per semester. However, there is a concern that additional credits may harm student performance. We find no evidence that increased course load harms performance, which supports the use of time-to-degree policy that uses course load as a lever.

What impact can universities actually have on course load? One avenue is to give students more options in course scheduling, such as evening and on-line courses (Witteveen and Attewell 2019). Advising can also play a prominent role in increasing credits per semester. Advising plans are at the heart of many of the finish-in-four initiatives on college campuses in the United States discussed previously (California State University 2019; University of Hawaii (Venit 2017); the University of South Dakota (<https://www.usd.edu/acpc/finish-in-four>); and Oklahoma State University (<https://cas.okstate.edu/advising/finish-in-four>)). However, advising on course load may need to be intensive; in a randomized experiment, an informational intervention designed to increase course load had no impact by itself (Huntington-Klein and Gill, 2019).

Course load policy does not need to be the only tool used for reducing time-to-degree. Increasing the number of summer classes is one way to increase the speed at which students accumulate credits. Policy on institutional barriers and resources, and student performance, is of course also important. Pike and Robbins (2020) find that the predominant factors affecting graduation rates are institutional characteristics that are either invariant or cannot be quickly or easily changed. Yet, they do find that per pupil expenditures for instruction are associated with better 4-year and 6-year graduation rates. Though not stated explicitly by the authors, factors such as peer tutoring and university efforts at remediation typically fall into this category. Walvoord and Pleitz (2016) provide non-causal evidence that peer tutoring is associated with higher first year grades. Additional support for the notion that advising, tutoring, and mentoring matter to degree completion is found in Deming and Walters (2017). The authors find both positive effects of school spending on degree completion and that spending on categories of academic support services such as advising and tutoring are particularly hard hit when there are state budget shocks.

Student support services and student-engagement programs have the potential to improve time to degree, particularly in the current COVID-19 environment. Support for students with disabilities may improve outcomes in light of some evidence that students with disabilities are less likely to complete a bachelor's degree if they enroll in college, controlling for pre-college characteristics (Lichtenberger 2010; Smalley et al. 2010). Based on survey results, Graham-Smith and Lafayette (2004) report that students with disabilities themselves highly value a caring staff that provides safety and security. Gaddy (2016) outlines strategies campus disability service providers can take to improve student performance, including the use of test-taking and writing strategies. As mentioned, students' mental and physical health may affect performance. Larson et al. (2016) conclude that support and assistance with students' health can improve outcomes, and that such assistance requires coordinated efforts across academic and student affairs (Larson et al. 2016, p. 600).

Evidence on the importance of student-engagement programs on retention and degree receipt is mixed. For example, Caviglia-Harris and Maier (2020) find that living-learning communities are positively associated with retention. In contrast, Culver and Bowman (2020) find no effect of first-year seminars on retention or four-year graduation. Johnson and Stage (2018) find little to no impact of so-called high-impact practices on 4-year or 6-year graduation rates at four-year public institutions.

Course load policy should be thought of as sitting alongside this array of interventions that focus on student performance, and requires similar institutional attention. Aforementioned advising efforts will be less effective if students are unable to take their preferred classes at the times needed to finish their course of study in a timely manner. This barrier to timely completion may be particularly relevant at a large public university such as ours. Course scarcity could cause students to take fewer classes, take the wrong classes, or change majors, all of which would increase time to degree (Kurlaender et al. 2014, pp. 33–35). Reporting on a survey of students and focus groups from three universities in the CSU system, Moore and Tan (2018) indicate that students themselves rate course availability as the number one obstacle to timely completion of their degrees. Among students who enrolled as freshmen, roughly half indicated that they, “often or very often have not been able to enroll in a class that they needed” (Moore and Tan 2018, p. 2).

It is important to note, however, that causal evidence that course scarcity affects time to degree is mixed. Kurlaender et al. (2014), using an IV strategy based on randomly assigned registration times, find no evidence of course scarcity on time to degree using data from the University of California, Santa Cruz. Robles et al. (2019) use a fuzzy regression discontinuity strategy based on waitlist cutoffs at De Anza Community College find that course scarcity increases the probability of dropout or transfer to another community college. These concerns about course availability reinforce the need for advising and mentoring as a way to assist students in timely degree completion and increase course load without being caught in bottlenecks, and also reinforce that policy pushing additional course load needs to first make sure that it's actually possible for students to take more courses at the institution.

Conclusion

In this paper we use administrative observational data in order to assess the causal effect of taking a full course load on student performance. We first focus on within-student variation in course load to avoid bias arising from student ability. We find no evidence of a negative effect of a full course load on student grades, and instead find a small positive effect.

Our baseline estimates are likely to be positively biased due to a relationship between course difficulty and course load, or the presence of time-varying outside work or family pressures. We find that controlling for course difficulty does not change results. We cannot measure outside pressures, but we provide four pieces of evidence supporting our conclusion. First, we note that, in the context studied, a significant determinant of course load is registration bottlenecks (Moore and Tan 2018), which would not bias results. Second, we control for lagged GPA, which is a proxy for time-varying academic performance as well as time-varying outside pressures under the assumption that those pressures are serially correlated, and find no difference in results. Third, we perform a simulation and find that outside pressures would need to have a correlation of around 0.399 with taking a full course load, and also a correlation of around 0.531 with grades, to generate the observed data if the true effect were meaningfully negative. Fourth, we show using Oster (2019) that obtaining a meaningfully large negative effect of course load on grades requires a fairly large (0.6) within-student R^2 for a model including both observable and unobservable controls. We then show using Cinelli & Hazlett (2020) that a meaningfully large negative effect requires an unobserved predictor to have variance orthogonal to the included controls that explains 17.4% or more of the within-student residual variance in both treatment and outcome.

There are two important takeaways from the evidence presented here. The first is that we find, to a reasonable degree of certainty, that there is no meaningfully large negative causal impact of increasing student course loads. Policy directives to improve four-year graduation rates by increasing course load are unlikely to have meaningful negative effects on student performance and learning. It is also unlikely that those average effects mask strong negative effects for weaker students, although it is possible that the 19.2% of students who always take the same number of classes would perform worse if pushed to do otherwise. As mentioned, our university policy encouraging students to take more credits was enacted several years after the first-time freshmen cohorts we study started their coursework. Future work looking at the causal effects of full course load as induced by policy would be useful. An analysis comparing the effects of course load for transfer students, as opposed to first-time freshmen, may also be a productive avenue for future research, but transfer students were not available in our data.

Our work contributes to the broader literature on college student performance. We contribute to a literature that has relatively few studies with plausibly causal estimates of the effect of increasing demands on students' time on their performance in other classes. A small body of associational literature will benefit from the addition of more plausibly causal results. The findings in this paper imply that one obvious solution—higher course loads—to a large and expensive nationwide problem—slow time-to-degree—does not appear to have the feared negative tradeoff for student performance.

Acknowledgements The authors are grateful for financial support from the California State University, Fullerton Graduation Initiative 2025.

Compliance with Ethical Standards

Conflict of interest The authors have no financial interests to declare.

References

- Adelman, C. (1999). Answers in the tool box. Academic intensity, attendance patterns, and bachelor's degree attainment. US Department of Education.
- Adelman, C. (2006) The toolbox revisited: Paths to degree completion from high school through college. US Department of Education.
- Attewell, P., Heil, S., & Reisel, L. (2012). What is academic momentum? And does it matter?'. *Educational Evaluation and Policy Analysis*, 34(1), 27–44. <https://doi.org/10.3102/0162373711421958>
- Attewell, P., & Monaghan, D. (2016). How many credits should an undergraduate take? *Research in Higher Education*, 57(6), 682–713. <https://doi.org/10.1007/s11162-015-9401-z>
- Babcock, P., & Marks, M. (2011). The falling time cost of college: Evidence from half a century of time use data. *The Review of Economics and Statistics*, 93(2), 468–478.
- Belfield, C., Jenkins, P. D., & Lahr, H. E. (2016). *Momentum: The academic and economic value of a 15-credit first-semester course load for college students in Tennessee*. 88. CCRC Working Paper.
- Berthelon, M., Bettinger, E., Kruger, D., & Montecinos-Pearce, A. (2019). The structure of peers: The impact of peer networks on academic achievement. *Research in Higher Education*, 60, 931–959.
- Bettinger, E. P., & Long, B. T. (2009). Addressing the needs of underprepared students in higher education does college remediation work? *Journal of Human Resources*, 44(3), 736–771.
- Bound, J., Lovenheim, M. F., & Turner, S. (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics*, 2(3), 129–157.
- Bound, J., Lovenheim, M. F., & Turner, S. E. (2012). Increasing time to baccalaureate degree in the United States. *Education Finance and Policy*, 7(4), 375–424.
- Brugiavini, A., Carraro, C., & Kovacic, M. (2020). Academic achievements: The effects of excess time to degree on GPA. *Education Economics*, 28(1), 46–66.
- California State University. (2019). Graduation Initiative 2025. <https://www2.calstate.edu/csu-system/why-the-csu-matters/graduation-initiative-2025>.
- Caviglia-Harris, J., & Maier, K. (2020). It's not all in their heads: The differing role of cognitive factors and non-cognitive traits in undergraduate success. *Education Economics*. <https://doi.org/10.1080/09645292.2020.1729702>
- Cinelli, C., & Hazlett, C. (2020). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1), 39–67.
- Cohn, E., Cohn, S., Balch, D. C., & Bradley, J., Jr. (2004). Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank. *Economics of Education Review*, 23(6), 577–586.
- College Board. (2018). Trends in higher education—The College Board. <https://trends.collegeboard.org/college-pricing/figures-tables/2017-18-state-tuition-and-fees-public-four-year-institutions-state-and-five-year-percentage>.
- Cornwell, C. M., Lee, K. H., & Mustard, D. B. (2005). Student responses to merit scholarship retention rules. *Journal of Human Resources XL*, 4, 895–917.
- Credé, M., Tynan, M., & Harms, P. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, 113(3), 492–511.
- Culver, K. C., & Bowman, N. A. (2020). Is What Glitters Really Gold? A Quasi-Experimental Study of First-Year Seminars and College Student Success. *Research in Higher Education*, 61, 167–196. <https://doi.org/10.1007/s11162-019-09558-8>
- Cyrenne, P., & Chan, A. (2012). High school grades and university performance: A case study. *Economics of Education Review*, 31(5), 524–542. <https://doi.org/10.1016/j.econedurev.2012.03.005>
- Daniłowicz-Gösele, K., Lerche, K., Meya, J., & Schwager, R. (2017). Determinants of students' success at University. *Education Economics*, 25(5), 513–532. <https://doi.org/10.1080/09645292.2017.1305329>
- Darolia, R. (2014). Working (and studying) day and night: Heterogeneous effects of working on the academic performance of full-time and part-time students. *Economics of Education Review*, 38, 38–50. <https://doi.org/10.1016/j.econedurev.2013.10.004>

- Deming, D., & Walters, C. (2017). *The impact of state budget cuts on U.S. postsecondary attainment*. Working Paper 23736, National Bureau of Economic Research.
- Denning, J. T. (2018). Born under a lucky star: Financial aid, college completion, labor supply, and credit constraints. *The Journal of Human Resources*, no. Forthcoming. <https://doi.org/10.3368/jhr.54.3.1116.8359R1>.
- Dynarski, S. (2008). Building the stock of college-educated labor. *Journal of Human Resources*, 43(3), 576–610.
- Ehrenberg, R. G., & Sherman, D. R. (1987). Employment while in college, academic achievement, and postcollege outcomes: A summary of results. *The Journal of Human Resources*, 22(1), 1–23.
- Emerson, J., Brooks, R. L., & McKenzie, E. C. (2009). College athletics and student achievement: The evidence at small colleges. *New Directions for Institutional Research*. <https://doi.org/10.1002/ir.314>
- Evans, W. N., Kearney, M. S., Perry, B. C., & Sullivan, J. X. (2017). *Increasing community college completion rates among low-income students: Evidence from a Randomized Controlled Trial evaluation of a case management intervention*. Working Paper 24150. National Bureau of Economic Research. <https://doi.org/https://doi.org/10.3386/w24150>.
- Fairlie, R. (2012). Academic achievement, technology and race: Experimental evidence. *Economics of Education Review*, 31(5), 663–679.
- Fosnacht, K., Copridge, K., & Sarraf, S. (2019). How valid is grit in the postsecondary context? A construct and concurrent validity analysis. *Research in Higher Education*, 60, 803–822.
- Fricke, H., Lechner, M., & Steinmayr, A. (2018). The effects of incentives to exercise on student performance in college. *Economics of Education Review*, 66, 14–39. <https://doi.org/10.1016/j.econedurev.2018.06.009>
- Gaddy, S. (2016). Consider additional support for your students through collaboration with your academic support area. *Disability Compliance for Higher Education*, 22(2), 5–5.
- Graham-Smith, S., & Lafayette, S. (2004). Quality disability support for promoting belonging and academic success within the college community. *College Student Journal*, 38(1), 90–99.
- Hanushek, E. (1979). Conceptual and empirical issues in the estimation of educational production functions. *The Journal of Human Resources*, 14(3), Summer, pp. 351–388.
- Hanushek, E. (2020). Education production functions. In S. Bradley & C. Green (Eds.), *Economics of education* (2nd ed., pp. 161–170). London: Academic Press.
- Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4), 451–464.
- Herzog, S. (2006). Estimating student retention and degree-completion time: Decision trees and neural networks vis-à-vis regression. *New Directions for Institutional Research*, 2006(131), 17–33.
- Horton, M. (2016). Why students should take at least 15 credits every semester. The California State University System News. California State University Office of the Chancellor. <https://www2.calstate.edu/csuo-system/news/Pages/Why-Students-Should-Take-15-or-More-Units-Every-Semester.aspx>.
- Huntington-Klein, N., & Gill, A. (2019). An informational intervention to increase semester credits in college. *Series of Unsurprising Results in Economics*, 1, 1–17.
- Ishitani, T. T. (2006). Studying attrition and degree completion behavior among first-generation college students in the United States. *The Journal of Higher Education*, 77(5), 861–885.
- Jackson, T., Weiss, K. E., Lundquist, J. J., & Hooper, D. (2003). The impact of hope, procrastination, and social activity on academic performance of MidWestern College students. *Education*, 124(2), 310.
- Johnson, S. R., & Stage, F. K. (2018). Academic engagement and student success: Do high-impact practices mean higher graduation rates? *Journal of Higher Education*, 89(5), 753–781. <https://doi.org/10.1080/00221546.1441107>
- Kalenkoski, C. M., & Pablonia, S. W. (2009). Does working while in high school reduce U.S. study time? *Social Indicators Research*, 93(1), 117–21.
- Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success*. NBER Working Paper No. 20749. (Revised September 2017).
- Knight, W. E. (2004). *“Time to Bachelor’s degree attainment: An application of descriptive, bivariate, and multiple regression techniques”* (Vol. 2). Association for Institutional Research: IR Applications.
- Kurlaender, M., Jackson, J., Howell, J. S., & Grodsky, E. (2014). College coursscarcity and time to degree. *Economics of Education Review*, 41, 24–39.
- Larson, M., Orr, M., & Warne, D. (2016). Using student health data to understand and promote academic success in higher education settings. *College Student Journal*, 50(4), 590–602.
- Lichtenberger, E. J. (2010). *A longitudinal study of Illinois High School Graduates with disabilities: A Six-Year Analysis of Postsecondary Enrollment and Completion (IERC 2010–4)IERC 2010–4*. Edwardsville, IL: Education Research Council.
- Machado, J. A. F., & Santos Silva, J. M. C. (2019). Quantiles via moments. *Journal of Econometrics*. <https://doi.org/10.1016/j.jeconom.2019.04.009>

- Martínez, I. M., Youssef-Morgan, C. M., Chambel, M. J., & Marques-Pinto, A. (2019). Antecedents of academic performance of university students: Academic engagement and psychological capital resources. *Educational Psychology*, 39(8), 1047–1067. <https://doi.org/10.1080/01443410.2019.1623382>
- Moore, C., & Tan, C. (2018). ‘Get Me from Point A to Point B:’ Student perspectives on barriers to timely graduation at the California State University. *Education Insights Center*.
- Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, 37(2), 187–204. <https://doi.org/10.1080/07350015.2016.1227711>
- Pike, G., & Robbins, K. (2020). Using panel data to identify the effects of institutional characteristics, cohort characteristics, and institutional actions on graduation rates. *Research in Higher Education*, 61, 485–509.
- Pu, S., Yan, Y., & Zhang, L. (2020). Peers, study effort, and academic performance in college education: Evidence from randomly assigned roommates in a flipped classroom. *Research in Higher Education*, 61, 248–269. <https://doi.org/10.1007/s11162-019-09571-x>
- Richardson, M., Abraham, C., & Bond, R. (2012). “Psychological Correlates of University Students’ Academic Performance: A systematic review and meta-analysis.” *Psychological Bulletin*, 138(2), 353–87.
- Robles, S., Gross, M., & Fairlie, R. W. (2019). *The effect of course shutouts on community college students: Evidence from Waitlist Cutoffs*. NBER Working Paper No. 26376.
- Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). New York: Springer.
- Rothstein, D. S. (2007). High school employment and youths’ academic achievement. *Journal of Human Resources* XLI, 1(1), 194–213. <https://doi.org/10.3368/jhr.XLII.1.194>
- Ruhm, C. J. (1997). Is high school employment consumption or investment? *Journal of Labor Economics*, 15(4), 735–776. <https://doi.org/10.1086/209844>
- Sanabria, T., Penner, A., & Domina, T. (2020). Failing at remediation? College remedial coursetaking, failure and long-term student outcomes. *Research in Higher Education*, 61, 459–484. <https://doi.org/10.1007/s11162-020-09590-z>
- Schmitt, N., Keeney, J., Oswald, F. L., Pleskac, T. J., Billington, A. Q., Sinha, R., & Zorzie, M. (2009). Prediction of 4-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology*, 94(6), 1479–1497.
- Scott-Clayton, J. (2011). On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, 46(3), 614–646. <https://doi.org/10.3368/jhr.46.3.614>
- Scott-Clayton, J., & Zafar, B. (2016). *Financial aid, debt management, and socioeconomic outcomes: Post-college effects of merit-based aid*. Working Paper 22574. National Bureau of Economic Research. <https://doi.org/https://doi.org/10.3386/w22574>
- Scrivener, S., Weiss, M.J., Ratledge, A., Rudd, T., Sommo, C., & Fresques, H. (2015). Doubling graduation rates: Three-year effects of CUNY’s accelerated study in associate programs (ASAP) for developmental education students. MDRC.
- Shapiro, D., A. Dundar, P.K. Wakhungu, X. Yuan, A. Nathan, and Y. Hwang. 2016. *Time to Degree: A National View of the Time Enrolled and Elapsed for Associate and Bachelor’s Degree Earners. (Signature Report No. 11)*. National Student Clearinghouse.
- Smalley, D. J., Lichtenberger, E. J., & Brown, K. S. (2010). *A Longitudinal Study of the Illinois High School Class of 2002: A Six-year Analysis of Postsecondary Enrollment and Completion (IERC 2010–3)*. Edwardsville, IL: Illinois Education Research Council.
- Stadler, M., Becker, N., Schult, J., Niepel, C., Spinath, F., Sparfeldt, J., & Greiff, S. (2018). The logic of success: The relation between complex problem-solving skills and university achievement. *Higher Education*, 76, 1–15. <https://doi.org/10.1007/s10734-017-0189-y>
- Stinebrickner, R., & Stinebrickner, T. R. (2004). Time-use and college outcomes. *Journal of Econometrics*, 121(1), 243–269. <https://doi.org/10.1016/j.jeconom.2003.10.013>
- Stinebrickner, R., & Stinebrickner, T. R. (2008). The causal effect of studying on academic performance. *The B.E. Journal of Economic Analysis & Policy*, 8(1), 1–53.
- Stinebrickner, R., & Stinebrickner, T. R. (2003). Working during school and academic performance. *Journal of Labor Economics*, 21(2), 473–491. <https://doi.org/10.1086/345565>
- Sweet, J., Swayze, S., & Busse, K. (2019). The relationship between psychological capital and GPA: A study of one freshmen cohort. *Journal of Higher Education Theory and Practice*, 19, 129–146.
- Szafran, R. F. (2001). The effect of academic load on success for new college students: Is lighter better? *Research in Higher Education*, 42(1), 27–50. <https://doi.org/10.1023/A:1018712527023>
- United States Census Bureau. (2020). Household pulse survey: Change in post-secondary education. <https://www.census.gov/data-tools/demo/hhp/#/?measures=CSR>.
- U.S. Department of Education. (2017). “Table 326.10.” Digest of Education Statistics 2016. 2017. https://nces.ed.gov/programs/digest/d16/tables/dt16_326.10.asp.
- Venit, Ed. (2017). Why even C students should consider taking 15 credits their first semester. *Student Success Insights Blog* (blog). 2017. <https://www.eab.com/blogs/student-success-insights/2017/08/15-to-finish>.

- Volkwein, J. F., & Lorang, W. G. (1996). Characteristics of extenders: full-time students who take light credit loads and graduate in more than four years. *Research in Higher Education*, 37(1), 43–68. <https://doi.org/10.1007/BF01680041>
- Wallis, A. L., Gretz, D. P., Rings, J. A., & Eberle, K. M. (2019). Assessing Marijuana use, anxiety, and academic performance among college students. *Journal of College Counseling*, 22, 125–137. <https://doi.org/10.1002/jocc.12125>
- Walvoord, M. E., & Pleitz, J. D. (2016). Applying matched sampling to evaluate a university tutoring program for first-year students. *Learning Assistance Review (TLAR)*, 21(1), 99–113.
- Witteveen, D., & Attewell, P. (2019). Delayed time-to-degree and post-college earnings. *Research in Higher Education*. <https://doi.org/10.1007/s11162-019-09582-8>
- Yue, H., & Fu, X. (2017). Rethinking graduation and time to degree: A fresh perspective. *Research in Higher Education*, 58, 184–213. <https://doi.org/10.1007/s11162-016-9420-4>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.