CALIFORNIA STATE UNIVERSITY
FULLERTON
MIHAYLO COLLEGE
OF BUSINESS AND ECONOMICS

Department of Economics
Working Paper Series

2018/009

Semester Course Load and Student Performance

Nick Huntington-Klein & Andrew Gill

August 2018

Semester Course Load and Student Performance

Nick Huntington-Klein and Andrew Gill

California State University, Fullerton

August 2018

Abstract: Many college students in the United States take longer than the proscribed four years to complete their bachelor's degrees. Long time-to-degree leads to billions of dollars of additional costs in higher education in the form of education costs and lost wages. Time-to-degree can be reduced if students to take more credits each term. However, an increased course load may lead students to reduce their time investment in each course, harming performance. Using longitudinal data on two cohorts of students at a regional four-year university with a high average time-to-degree, we fail to find any evidence that a high course load has a negative impact on student performance in class. This result is consistent with a model where students substitute time away from non-education activities when their course loads increase.

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

I. Introduction

Increased time-to-degree completion 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 major public-policy concerns in higher education. Among first-time full-time students seeking a bachelor's degree and commencing their studies in 2009, only 39.8 percent graduated from their first institution attended within 4 years (U.S. Department of Education 2017). The problem is especially severe at public institutions, where four-year graduation rates for the 2009-entering cohort stood at 34.8 percent (58.6 percent at six years), while the rate among private not-for-profit schools was 53 percent (65.6 percent at six years).

Low four-year graduation rates are reflected in 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 percent were enrolled for 6 years and 18.2 percent 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. Given an average tuition outlay at public institutions of $9,970 (College Board 2018) and an average $23,081 per year in lost earnings, the extra time to completion beyond 4 years amounts to about $41.7 billion in additional costs to students receiving a bachelor's degree per cohort.[1] This calculation does not include the expenditures made by government on education.

Common explanations for low graduation rates and increased time-to-degree completion typically center on the relative importance of factors such as student preparedness, 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). Regarding financial need, for example, a portion of this

[1] Lost earnings calculated by comparing wages of current college enrollees aged 20-25 against current BA holders aged 20-25 in the 2016 Current Population Survey.

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; Angrist, Lang, and Oreopoulos 2009; Scott-Clayton 2011). Students themselves focus on slightly different concerns regarding increased time-to-degree completion. Moore and Tan (2018) analyzed student perceptions on time-to-degree and found that course availability and work and family obligations serve as 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). Almost by necessity, a reduction in time-to-degree will require an increase in the number of credits each student completes each term. Completing a typical 120- credit degree program in 4 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).

An obvious policy prescription that follows from the available evidence is that universities could reduce time-to-degree completion by encouraging students to take more credits (Attewell and Monaghan 2016). Yet, equally obvious are the potential unintended negative consequences that could follow, as taking additional credits has the potential to crowd out study time per course. A policy encouraging students to take more credits may be ill advised 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. If, on the other hand, there is no discernable effect of credits attempted on student performance, a focused approach to advising students to take more credits could be a cost effective method for improving time-to-degree completion.

The question of whether increased class loads harms student performance takes on greater importance in light of evidence indicating that students themselves believe this to be the case. Volkwein and Lorang (1996) find that students who complete multiple semesters with fewer than 15 credits, so called "extenders," do so in part because of concerns about maintaining a high GPA. Cornwell, Lee, and

Mustard (2005) find that freshmen resident students in Georgia responded to GPA requirements for maintaining HOPE merit scholarships by reducing the likelihood of full-time enrollment, increasing course withdrawals, and diverting effort to summer classes to reduce course difficulty.

Given the importance of understanding the relationship between student performance and course loads, we investigate the causal effect of course load on grades using a rich set of administrative data from a four-year public university in California at which the four-year graduation rate is low despite a six-year graduation rate above the national average, indicating long time-to-degree. We follow 8,015 freshmen students starting in Fall 2010 or Fall 2011. Our particular interest is in analyzing the effect of course loads on student performance among students enrolled full-time.

Assessing whether course loads affect student performance is complicated by unobserved student characteristics as well as the endogenous nature of the choice to take 12 or 15 credits. In one empirical strategy, we estimate standard fixed-effects regressions to isolate within-student variation and compare students' grades in semesters when they take 12 credits against grades in semesters when they take 15. 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 0.005 standard deviation increase in course GPA, rather than a negative effect of any size. When we limit the analysis to students who graduated during our sample window, the effect becomes smaller (0.002) and statistically insignificant, but still is not negative. In a second empirical strategy, we estimate the effect of course load on students' grades using Coarsened Exact Matching (CEM) estimators (Iacus et al., 2012). We again find no negative effects of course load on student performance. Our results are robust to including time-varying student performance measures in the empirical analysis and accounting for the difficulty of the course mix chosen. Overall, our results support a conclusion that encouraging students 15 credits may be an effective method to decrease time to degree completion without harming student performance.

II. Literature

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, this implies a negative effect of course load on performance. Stinebrickner and Stinebrickner (2004), for example, find a strong positive association between student study time and first-year grades. Stinebrickner and Stinebrickner (2008), in turn, find a 0.36 increase in GPA for an additional study hour per day, where the effect of study time on grades was identified using the availability of video games via one's roommate.

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. Kalenkoski and Pabilonia (2009) find that working while in school reduces study time by only a small degree for high school students. Babcock and Marks (2011) examine study time in college and find little evidence that employment crowds out study time.

Weak effects of employment on actual study time are accompanied by inconsistent evidence of the effects of employment on performance. Ruhm (1997) reports on the general lack of consensus in early work on student employment and school performance. Oettinger (1999) finds fairly large negative effects of high levels of employment during the academic year on high school grades for black and Hispanic students, but not for whites. Rothstein (2007) finds very small to zero effects of hours worked during the school week in high school on GPA, with the smallest effects resulting when student level fixed effects are considered and an instrument is used to isolate the effect of work time.

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. More recently, Stinebrickner and Stinebrickner (2003), using an identification strategy that exploits available hours of work from "assigned" required jobs in a small liberal arts college, show that increasing hours of work by one hour per week reduces semester GPA by 0.162. Darolia (2014), in contrast, using data from the National Longitudinal Survey of Youth 1997 (NLSY97), finds no evidence that students' grades are harmed by working, but does find that credits completed for are reduced for four-year full-time students.

Another possible avenue where time commitments could crowd out study time and reduce grades are in sports and other extracurricular activities. In this regard, Aries et al. (2004) find no evidence that participation in athletic or non-athletic extracurricular activities affected final grades. Emerson et al. (2009) find that recruited college athletes underperform relative to students who are not athletes, but suggest that confounders other than time commitments drive these results.

In short, the evidence is generally consistent with a potential negative impact of additional time demands on learning, but only in some situations. The difference may be due to whether these demands are strong enough to actually reduce study time. It may be that students who take on additional responsibilities, at least at low levels, are able to substitute from other tasks besides study time.

The second theoretical framework, which implies a positive relationship between course load and student performance, comes generally from outside of the economics literature and focuses on the level of student involvement in their academic lives. 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 with an elasticity above 1. This is the concept of "academic momentum," a term first discussed by Adelman (1999, 2006), which suggests, among other things, that students who complete more credits in their first year at college are more likely to obtain a degree. Attewell, Heil, and Reisel (2012) refine the concept of academic momentum by outlining possible mechanisms. Most notably 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).

The available evidence on academic momentum, typically identified by matching estimators, broadly supports a positive association between the completion of first-semester or first-year credits and student success. Attewell, Heil, and Reisel (2012) find that enrolling part time in the first semester is associated with a decreased probability of degree completion for students at two and four-year colleges.

Attewell and Monaghan (2016) find effects on a similar scale, and that the effects are stronger among minority students. Belfield, Jenkins, and Lahr (2016) additionally find that the effects of higher course load are stronger if maintained over several terms. Increased course loads are also found to improve completion times. Volkwein and Lorang (1996), Knight (2004), Belfield, Jenkins, and Lahr (2016), and Venit (2017) all find direct or indirect evidence that increase course load reduces completion time., although Cornwell, Lee, and Mustard (2005) finds only mixed evidence.

Several successful large-scale policy implementations, intended to improve completion by a number of simultaneous interventions, include increased course loads. Scott-Clayton (2011) finds that the PROMISE merit-based scholarship program in West Virginia, which implemented a course load minimum, improved both earned credits by the end of freshman year and BA completion 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.

The bulk of the literature informing the question of whether increased course loads 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. We turn next to the small amount of direct evidence on whether taking increased course loads impairs student performance in the areas of retention and course grades. Szafran (2001) reports a non-causal association between first-semester course load and first-semester retention, finding that the improved retention operates mainly through improved grades. Jackson et al. (2003) similarly find a positive correlation between GPA and the number of units completed. Venit (2017), reporting on the University of Hawaii's "15 to Finish" program, indicates that the university's analysis of historical records found no evidence that student performance was harmed when taking a 15 credit course loads. Further, based on data from the Student Success Collaborative, Venit reports that students who take at least 15 credits per term in their first year were more likely to persist in college and achieved higher GPAs than students taking fewer that 15 credits. In an effort to control for the confounding affect of student ability, the authors conducted the same analysis over four levels of high-school GPA. They found that even the lowest performing high-school students benefit from taking at least 15 credits.

The evidence on increase 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, 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.

III. Data

We use administrative data from a major four-year university, provided by the office of Institutional Research & Analytical Studies (IRAS). Data are at the student-course level and include information on all courses attempted and grades received for two incoming cohorts of freshmen students, 3,874 students beginning Fall 2010 and 4,141 beginning Fall 2011.[2]

We observe student course-taking and grades through the end of the Spring 2017 semester. We observe courses and grades for 18,562 courses while the students were freshmen, 13,944 courses while the students were sophomores, 12,942 courses while the students were juniors, and 14,479 courses while the students were seniors. The drop in courses for students in higher class standing is due to dropout, followed by an increase for seniors attributable to students who spend multiple years as seniors. There were 34.6 classes attempted, on average, across all students. 64% of the students from the incoming cohorts had graduated by Spring 2017. Only 20.7% graduated within four years. We drop all courses taken during summer terms, or during terms where the student was part-time.

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

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 or SAT scores. In addition to a student's own background and performance, we observe 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. As shown, a large fraction of the students in the incoming cohorts analyzed report that they are Hispanic (0.403) and a large fraction report receiving some financial aid (0.809). In this sample, receiving no financial aid is a strong indicator of being an international student. On average, students attempted 4.63 classes per semester, where taking 5 classes is a full course load and 12 is the minimum necessary to be considered a full-time student. A revealing feature of the descriptive data on classes attempted is that the modal number of classes over the time period analyzed was 4 for 32.8 percent of students in the sample. Figure 1 shows that across all students and semesters, the modal number of classes attempted was 4, with 5 close behind.

Contrary to what might be expected, while the decision to take 4 or 5 courses certainly 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. Excluding students who dropped out after one semester, in fully 48.8% of all semesters, students take a number of courses that is not equal to their personal modal number of classes. Even if reduced-credit courses are ignored so that the options for the number of courses taken is binary – "fewer than 15 credits" vs. "15 credits or more" – 27.1 of all semesters are still completed with the non-modal number of classes. 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 overenrolled 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.

IV. 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 observed relationship between classes taken and grades that is non-causal. 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 positive returns to scale to the production of good grades (for example, if by taking many classes in the same semester one saves on transportation and task-transition costs), 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 (positive) consumption shock requiring them to work more (less), they may be likely to reduce (increase) 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 (many) classes, they can only reduce (increase) effort rather than taking fewer (more) 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 to maintain a full course load and more motivated students may be more likely 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 number of courses taken 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 GPA_{ijt} 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}):

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

We move β_{1j} to the left-hand side of Equation (1), using average grades earned by students in the course as a proxy, and replacing the left-hand side with grades standardized within course:

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

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 (ν_{it}).

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

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 the relevant policy would not attempt to change the number of classes while holding course mix constant. 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 ν_{it}, which is likely nonzero because pressures on work hours are likely to be in both ε_{it} and ν_{it}. Basic estimates will be biased by this correlation, and we estimate partial identification bounds on $\hat{\beta}_2$ to determine the strength of the relationship necessary to change the substantive results of the paper.

V. 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 main findings. Next, we provide some supplementary results examining the relationship between classes attempted, persistence, and completion. In a final section we report our Coarsened Exact Matching (CEM) estimates of the effect of course load on student performance.

V.1 Fixed Effects results

Our fixed effects estimates explaining student performance appear in Table 2. The dependent variable for all regressions in Table 2 is a student's GPA in each class, standardized within the particular class section. In Column 1 we show pooled OLS results giving the relationship between the number of classes attempted in the term and standardized GPA, controlling for a student's class standing. In Column 2 we add student level characteristics. The within estimates regressing classes attempted on standardized GPA appear in Columns 3 and 4 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, during the period covered by our data.

As shown in Column 1, an additional class attempted is associated with a 0.041 standard deviation increase in GPA. However, this estimate does not account for student ability. We show in Column 2 that the association between classes attempted and GPA is reduced by roughly 29 percent when we add student-level characteristics. Controlling for GPA and admissions index, all race/ethnic groups show lower standardized GPA than whites (the reference category) with otherwise similar characteristics, and a one point increase in high school GPA is associated with a 0.165 standard deviation increase in college course GPA. It is of some interest that the admissions index is not associated with college GPA once we control for high school GPA, although the coefficient for the admissions index, even more so than GPA or SAT, is subject to range restriction.

Column 3 shows a within-student estimate of the effect of classes attempted on GPA. The effect is positive, but is so small (0.005) that it is of little practical significance. Importantly, though, we can reject to a high degree of precision that the effect of a given average student taking a full course load is negative. Additionally, Column 3 gives some insight to the likely correlation between fixed unobserved student heterogeneity and classes attempted. The implied effect of classes attempted is reduced significantly moving from the pooled OLS results in Column 1 or the pooled OLS results controlling for student characteristics in Column 2 to the within-student estimates reported in Column 3, suggesting 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 show that there is no positive effect of classes attempted for students who graduate within the sample window. Instead, the effect is concentrated among non-graduates (Column 5). The effect is not large for either subgroup, but is more positive for non-graduates, who are less likely to choose higher course loads in the first place.

[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).

V.2 Observed Sources of Bias

Student ability is one determinant of course load, but there may be other important factors. Table 3 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. As shown in Column 1 of Table 3, students taking more classes tend to be, unsurprisingly, students with higher high school GPAs. Regarding student demographic characteristics, controlling for GPA, there is no gender difference in the tendency to take more courses, but there are racial and ethnic differences, with black students taking the most, followed by white students. Hispanic students, nearly 40% of the sample, take on average .16 fewer courses per term than whites.

The other columns in Table 3 show results giving the influence of time-varying student characteristics on the number of classes taken, and include student fixed effects. As shown in Column 2, students tend to take more courses when the courses they are taking are easier. A student taking a mix of courses for which the average grade given among all students is one point higher will take 0.079 more courses. Column 3 shows the results when we adjust the average grade in the course for the demographic characteristics and previous college GPA of the other students who typically take that course. An average course grade one GPA unit higher that could be expected given who takes the course is associated with the student taking 0.236 more courses. Columns 4 and 5, which omit the student’s first term (and first two terms, respectively) examine the student’s recent performance in college. Keeping in mind that these regressions include student fixed effects, the results in Columns 4 and 5 show that a student who has recently been doing better than they usually do is likely to take more courses, which could be interpreted as confidence, or learning one’s limits.

Given that we find significant predictors of the number of courses taken in Table 3, Table 4 displays results when we examine whether any of these predictors can explain the positive relationship between classes taken and grades. Columns 1 and 2 display the results when we add the average course grades. While the left-hand side variable is already adjusted for course difficulty, it is still possible that average course grades could act as a confounder, because taking multiple easy classes at once may allow a student to over-perform in all of them at once. However, neither addition eliminates the positive relationship between classes taken and GPA, nor does controlling for recent or prior performance in Columns 3 and 4. While the relationship between classes and grades shrinks here, it remains positive, indicating that the positive relationship is not a result of students, for example, taking more classes as they discover courses for which courses they have a comparative advantage. Finally, in Column 5, we include fixed effects for declared major to account for differing institutional standards for how many courses students should take, especially in STEM. The effect of classes attempted on course grades remains positive.

The relationship between courses taken and grades remains positive and statistically significant. While the effect is small, the qualitative result of importance is that we find 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.

V.3 Unobserved Work Requirement as a Source of Bias

Adding controls for confounders does not reverse the original prediction. Even including all listed confounders at once leaves a positive coefficient of .007 ($p < .001$) on the classes attempted variable. However, there is an important confounder missing from our analysis – that of consumption shocks or time-varying work pressures. At the setting studied, significant portions of students work either part or full time while attending classes. Theoretically, we would expect that changes in courses taken driven by factors that increase work hours would reduce grades and courses taken at the same time, leading to a positive bias.

Since we do not observe work hours, we perform a simulated sensitivity test, inspired by Rosenbaum (2002) bounds, but applied outside the context of propensity score matching. 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

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

where $U_{it} \sim Unif[0,1]$. 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 $\{0,0.1,0.2,\ldots,1\}$, which in effect generates correlations between Z_{it} and GPA_{it}^{STD} of between 0 and .43, and correlations between Z_{it} and $Class_{it}$ of between 0 and .82. 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 this process 100 times for each combination of δ_1 and δ_2, and for each of four models: the baseline (Table 2 Column 3), controlling for average grade in the class (Table 4 Column 1), controlling for student's recent performance (Table 4 Column 4), 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 1 shows the minimum correlations necessary to generate a statistically significant negative coefficient on $Class_{it}$ at the 95% level (i.e. 95 or more out of 100 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. Figure 1 suggests that such an omitted predictor would need to have a correlation of about .2 or better with both $Class_{it}$ and GPA_{it}^{STD}, which is relatively strong. If the omitted predictor has a correlation of less than .2 with GPA_{it}^{STD}, it is possible to still produce a negative result if the correlation with $Class_{it}$ is strong enough, but the opposite does not appear to be true - if the correlation between Z_{it} and $Class_{it}$ is less than about .15, very high correlations between Z_{it} and GPA_{it}^{STD} will still not produce a negative significant result.

So, a relatively strong correlation between an omitted predictor and $Class_{it}$ is necessary to produce a negative result, but a correlation of .2 is certainly not unheard of. However, even in these statistically significant cases, the negative relationship between $Class_{it}$ and GPA_{it}^{STD} is still not large. The coefficient on $Class_{it}$ along Figure 1's lines ranges from -0.001 to -0.005. Considering only statistically significant negative coefficients in the simulation, the average coefficient on $Class_{it}$ was -0.03, and in no simulation for any model is the coefficient lower than -0.086 (noting that stronger correlations, and thus more negative coefficients, could be induced by increasing δ_1 or δ_2 past 1). We conclude that, even with the likely presence of positive omitted variable bias in our main results, adjusting for this bias is unlikely to lead to a meaningfully large negative relationship between $Class_{it}$ and GPA_{it}.

V.4 Supplementary Results

In Table 5 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 in the term they are taken, may lead to burnout so that students are less likely to return. We do not find this here. Instead, there is a modest positive relationship between taking more classes and persisting to the next term, consistent with much of the literature discussed in Section II.

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 $Class_{it}$ is an average over all enrolled terms. Without within-student variation, these estimates are to be considered non-causal. There is a very strong relationship between taking more classes and graduating; students who take one more class per term graduate almost 20 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 three quarters of a year less time to graduate. This three quarters of a year measure is, notably, less than the full year decrease one might expect mechanically, given that attaining 120 credits with four classes per term should take five years, and with five classes per term should take four years.

V.5 Coarsened Exact Matching Estimates

In this section we repeat the analysis using a matched sample so as to focus identification on the most closely comparable groups. This approach provides treatment-on-treated estimates, and allows us to provide an analysis in which the treatment and control groups are as closely comparable as possible. We provide matching estimates in two ways, first providing a baseline matching estimator that uses pre-college student characteristics. Because these characteristics are all fixed, this analysis does not provide much beyond our fixed effects estimate, and we provide it largely for comparison. Our matching estimate of focus is dynamic, in which we match students on a term-by-term basis using their pre-college characteristics as well as their recent academic performance. This uses information about the source of some of the within-student variation in course load to account for selection in a way that the fixed effects estimator does not provide. If the dynamic matching estimator differs, this suggests that our initial results may be heavily influenced by dynamic unobservable selection pressures.

We apply a Coarsened Exact Matching (CEM) estimator, which is not as model-dependent as other matching estimators, and emphasizes comparability of subjects over reduction in variance (Iacus et al. 2012). CEM takes every variable used in the empirical analysis and coarsens any continuous variables into bins. Strata are determined by which bin a given observation falls in, and observations are only matched if there are other observations in the exact same strata across all variables.

In the first case, matching is at the student-term level using only measures that one could observe before students started college. These measures are gender, financial aid, race, cohort, initial major STEM vs. non-STEM, parental education, and high-school GPA. We obtained the matching set based on one observation per student, matching students who took 4 classes their first term with students who took 5 classes their first term. Course load was coarsened into “5 classes or more” vs. “fewer than 5 classes.” We then used the same matching set for all periods. In all, 99.8 percent of observations were matched.

The top panel of Table 6 provides estimates that are highly similar to the fixed effects estimates from earlier sections. This suggests a comparability between the fixed effects and matching estimates, and is a base for comparison as we introduce dynamic matching.

In a second approach to matching, we performed the matching separately in each semester. We match to students in the same semester and student standing, on all student background characteristics used in the first approach, on the number classes a student took last term (5 and above or below 5), and the average GPA a student earned last term. Note that this procedure excludes the first student term since we do not observe a lagged GPA or class numbers. In all, 88.3 percent of observations were matched. To assess the effect of classes attempted on performance, we use fixed effects on the matched subsample. The results are reported in Columns 5 and 6 of Table 6. As shown, we continue to find very small positive effects of classes attempted on performance.

VI. Conclusion

In this paper we use administrative observational data in order to assess the causal effects of increasing student course load on student performance. We first focus on within-student variation in course load to avoid bias arising from student ability using a standard fixed-effects methods. We find no evidence of a negative effect of higher course load on student grades, and instead find a small positive effect. We perform robustness and bounding tests to ensure that our estimates are not biased by endogenous factors such as course difficulty or student work pressures, and our results stand. Work pressures would need to be a very strong source of bias in order for the true effect of increased course load to be a meaningfully large negative number. In a second approach, we apply fixed effects methods to subsamples from Coarsened Exact Matching. We again find small positive effects of course load on student performance.

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 negative causal impact of increasing student course loads. There is a very strong overall positive correlation between course load and performance in class. Even when we adjust this relationship for student ability by focusing on within-student variation, we find a small positive effect of higher course load, especially within the group of students who do not graduate, who are likely weaker overall. Policy directives to improve four-year graduation rates by increasing course load are unlikely to have meaningful negative effects on student performance and learning.

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. We especially contribute to the understanding of the performance implications of increasing a student's level of interaction with their college environment by having them take more courses. A small body of associational literature will benefit from the addition of more plausibly causal results. And, further, our work implies 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.

VII References

Adelman, Clifford. 1999. “Answers in the Tool Box. Academic Intensity, Attendance Patterns, and Bachelor’s Degree Attainment.” US Department of Education.

———. 2006. “The Toolbox Revisited: Paths to Degree Completion From High School Through College.” US Department of Education.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. “Incentives and Services for College Achievement: Evidence from a Randomized Trial.” *American Economic Journal: Applied Economics* 1 (1): 136–63. https://doi.org/10.1257/app.1.1.136.

Aries, Elizabeth, Danielle McCarthy, Peter Salovey, and Mahzarin R. Banaji. 2004. “A Comparison of Athletes and Non-Athletes at Highly Selective Colleges: Academic Performance and Personal Development.” *Research in Higher Education* 45 (6): 577–602. https://doi.org/10.1023/B:RIHE.0000040264.76846.e9.

Attewell, Paul, Scott Heil, and Liza Reisel. 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, Paul, and David Monaghan. 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, Philip, and Mindy Marks. 2011. “The Falling Time Cost of College: Evidence from Half a Century of Time Use Data.” *The Review of Economics and Statistics* 93 (2): 468–78. https://doi.org/10.1162/REST_a_00093.

Belfield, Clive, Paul Davis Jenkins, and Hana Elizabeth Lahr. 2016. “Momentum: The Academic and Economic Value of a 15-Credit First-Semester Course Load for College Students in Tennessee.” 88. CCRC Working Paper.

Bettinger, Eric P., and Bridget Terry Long. 2009. “Addressing the Needs of Underprepared Students in Higher Education Does College Remediation Work?” *Journal of Human Resources* 44 (3): 736–71. https://doi.org/10.3368/jhr.44.3.736.

Bound, John, Michael F. Lovenheim, and Sarah Turner. 2010. “Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources.” *American Economic Journal: Applied Economics* 2 (3): 129–57. https://doi.org/10.1257/app.2.3.129.

Bound, John, Michael F. Lovenheim, and Sarah E. Turner. 2012. “Increasing Time to Baccalaureate Degree in the United States.” *Education Finance and Policy* 7 (4): 375–424.

College Board. 2018. “Trends in Higher Education - The College Board.” 2018. https://trends.collegeboard.org/college-pricing/figures-tables/2017-18-state-tuition-and-fees-public-four-year-institutions-state-and-five-year-percentage.

Cornwell, Christopher M., Kyung Hee Lee, and David B. Mustard. 2005. “Student Responses to Merit Scholarship Retention Rules.” *Journal of Human Resources* XL (4): 895–917. https://doi.org/10.3368/jhr.XL.4.895.

Darolia, Rajeev. 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.

Dynarski, Susan. 2008. “Building the Stock of College-Educated Labor.” *Journal of Human Resources* 43 (3): 576–610. https://doi.org/10.3368/jhr.43.3.576.

Ehrenberg, Ronald G., and Daniel R. Sherman. 1987. “Employment While in College, Academic Achievement, and Postcollege Outcomes: A Summary of Results.” *The Journal of Human Resources* 22 (1): 1–23. https://doi.org/10.2307/145864.

Emerson, John, Rachelle L. Brooks, and Elaine Croft McKenzie. 2009. “College Athletics and Student Achievement: The Evidence at Small Colleges.” *New Directions for Institutional Research* 2009 (144): 65–76. https://doi.org/10.1002/ir.314.

Evans, William N., Melissa S. Kearney, Brendan C. Perry, and James X. Sullivan. 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/10.3386/w24150.

Gibbons, Charles E., Juan Carlos Suárez Serrato, and Michael B. Urbancic. "Broken or Fixed Effects?," *Journal of Econometric Methods* (Forthcoming).

Herzog, Serge. 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. https://doi.org/10.1002/ir.185.

Iacus, Stefano, Gary King, and Giuseppe Porro. 2012. "Causal Inference without Balance Checking: Coarsened Exact Matching." *Policy Analysis* 20: 1-24

Ishitani, Terry T. 2006. “Studying Attrition and Degree Completion Behavior among First-Generation College Students in the United States.” *The Journal of Higher Education* 77 (5): 861–85.

Jackson, Todd, Karen E. Weiss, Jesse J. Lundquist, and Denise Hooper. 2003. “The Impact of Hope, Procrastination, and Social Activity on Academic Performance of MidWestern College Students.” *Education* 124 (2): 310.

Kalenkoski, Charlene Marie, and Sabrina Wulff Pabilonia. 2009. “Does Working While in High School Reduce U.S. Study Time?” *Social Indicators Research* 93 (1): 117–21. https://doi.org/10.1007/s11205-008-9412-7.

Knight, William E. 2004. “Time to Bachelor’s Degree Attainment: An Application of Descriptive, Bivariate, and Multiple Regression Techniques.” Volume 2. IR Applications. Association for Institutional Research.

Moore, Colleen, and Connie Tan. 2018. “‘Get Me from Point A to Point B:’ Student Perspectives on Barriers to Timely Graduation at the California State University.” Education Insights Center.

OECD. 2017. “Education at a Glance 2017: OECD Indicators.” OECD Publishing.

Oettinger, Gerald S. 1999. “Does High School Employment Affect High School Academic Performance?” *ILR Review* 53 (1): 136–51. https://doi.org/10.1177/001979399905300108.

Rosenbaum, Paul R. 2002. *Observational Studies*. 2nd ed. New York: Springer-Verlag.

Rothstein, Donna S. 2007. “High School Employment and Youths’ Academic Achievement.” *Journal of Human Resources* XLII (1): 194–213. https://doi.org/10.3368/jhr.XLII.1.194.

Ruhm, Christopher J. 1997. “Is High School Employment Consumption or Investment?” *Journal of Labor Economics* 15 (4): 735–76. https://doi.org/10.1086/209844.

Scott-Clayton, Judith. 2011. “On Money and Motivation: A Quasi-Experimental Analysis of Financial Incentives for College Achievement.” *Journal of Human Resources* 46 (3): 614–46. https://doi.org/10.3368/jhr.46.3.614.

Scrivener, Susan, Michael J. Weiss, Alyssa Ratledge, Timothy Rudd, Colleen Sommo, and Hannah Fresques. 2015. “Doubling Graduation Rates: Three-Year Effects of CUNY’s Accelerated Study in Associate Programs (ASAP) for Developmental Education Students.” MDRC.

Shapiro, Doug, Afet Dundar, Phoebe Khasiala Wakhungu, Xin Yuan, Angel Nathan, and Youngsik 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.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2004. “Time-Use and College Outcomes.” *Journal of Econometrics* 121 (1): 243–69. https://doi.org/10.1016/j.jeconom.2003.10.013.

———. 2008. “The Causal Effect of Studying on Academic Performance.” *The B.E. Journal of Economic Analysis & Policy* 8 (1): 1–53. https://doi.org/10.2202/1935-1682.1868.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2003. “Working during School and Academic Performance.” *Journal of Labor Economics* 21 (2): 473–91. https://doi.org/10.1086/345565.

Szafran, Robert 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.

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, James Fredericks, and Wendell G. Lorang. 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.

Table 1: Descriptive Statistics

	Mean	(SD)
Student-level data:		
Female	0.576	
Received Financial Aid	0.809	
Admissions Index	3556.971	(650.069)
High school GPA	3.323	(0.377)
Modal number of classes/term is 4	0.328	
Modal number of classes/term is 5	0.355	
White	0.253	
Black	0.026	
Hispanic	0.403	
Asian / Pacific Islander	0.224	
Other / Mixed	0.095	
Other levels:		
GPA in Class	2.811	(0.783)
Classes attempted in term	4.63	(0.561)
Non-modal number of units taken this term	0.488	

Figure 1: Distribution of the Number of Classes Attempted per Term

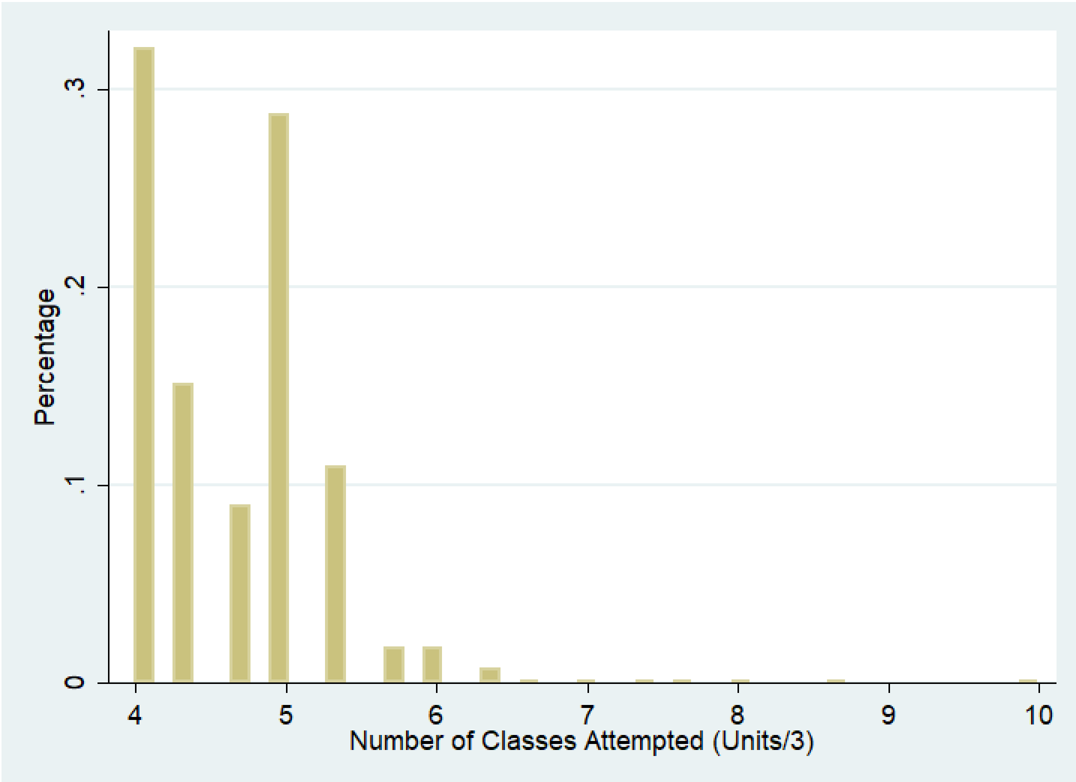


Table 2: Classes Attempted and Class Performance

	Dependent Variable: Standardized Class GPA				
	(1)	(2)	(3)	(4) Graduates Only	(5) Non-Grads Only
Classes Attempted This Term	0.041*** (0.002)	0.029*** (0.002)	0.005*** (0.002)	0.002 (0.002)	.008* (0.005)
Female		0.002 (0.002)			
Black		-0.066*** (0.007)			
Hispanic		-0.052*** (0.003)			
Asian / PI		-0.009*** (0.003)			
Other / Mixed		-0.024*** (0.004)			
Admission Index (Standardized)		0.006*** (0.001)			
Financial Aid		-0.008*** (0.003)			
High School GPA		0.165*** (0.003)			
Student Standing Fixed Effects	Y	Y	Y	Y	Y
Student Fixed Effects	N	N	Y	Y	Y
N Observations	59,668	59,390	59,668	44,952	14,716
N Students			7,995	5,132	2,863

Table 3: Predicting the Number of Classes Taken

	Dependent Variable: Number of Classes Taken this Term (Units/3)				
	(1)	(2)	(3)	(4)	(5)
Female	-0.001				
	(0.005)				
Black	0.026*				
	(0.015)				
Hispanic	-0.160***				
	(0.006)				
Asian / PI	-0.055***				
	(0.007)				
Other / Mixed	-0.028***				
	(0.009)				
Admission Index (Standardized)	0.011***				
	(0.003)				
Financial Aid	-0.003				
	(0.006)				
High School GPA	0.087***				
	(0.007)				
Average Grade in Course		0.079***			
		(0.011)			
Population-Adjusted Average Grade in Course			0.236***		
			(0.024)		
GPA Last Term				0.082***	0.078***
				(0.005)	(0.005)
Cumul. GPA Last Term					0.150***
					(0.012)
Student Standing Fixed Effects	Y	Y	Y	Y	Y
Student Fixed Effects	N	Y	Y	Y	Y
N Observations	59,647	59,927	59,927	44,555	37,046
N Students		8,015	8,015	7,666	6,986

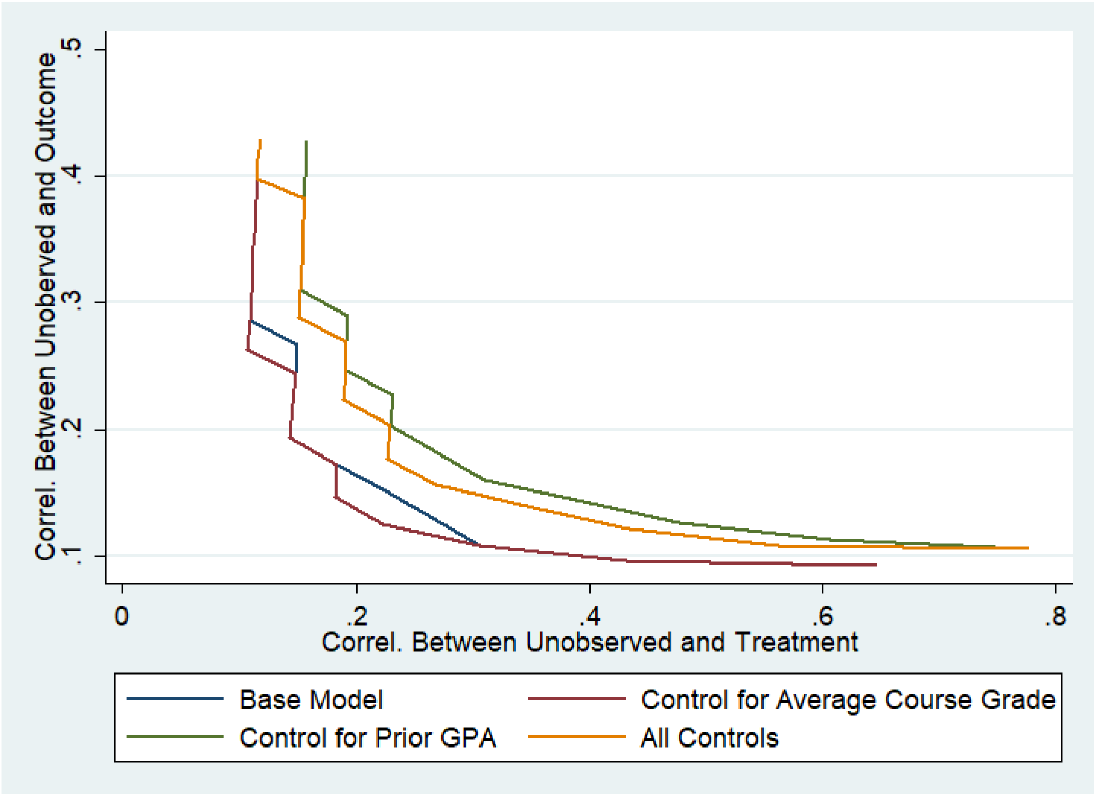
Table 4: Robustness of the Classes Attempted and Class Performance Link

	Dependent Variable: Standardized Class GPA				
	(1)	(2)	(3)	(4)	(5)
Classes Attempted This Term	0.016***	0.017***	0.006***	0.008***	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Average Grade in Course	0.019***				
	(0.004)				
Population-Adjusted Average Grade in Course		-0.044***			
		(0.009)			
GPA Last Term			0.099***	0.064***	
			(0.005)	(0.006)	
Cumul. GPA Last Term				-0.237***	
				(0.013)	
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	59,668	59,668	44,405	36,930	59,668
N Students	7,995	7,995	7,645	6,973	7,995

Table 5: 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)
Classes Attempted This Term	0.009***	0.018***		
	(0.002)	(0.004)		
Average Classes Attempted			0.198***	-0.787***
			(0.014)	(0.027)
Female	-0.001		0.051***	-0.211***
	(0.002)		(0.010)	(0.018)
Black	0.001		-0.056*	0.129**
	(0.005)		(0.033)	(0.059)
Hispanic	0.010***		0.009	0.076***
	(0.002)		(0.013)	(0.022)
Asian / PI	0.013***		0.041***	0.114***
	(0.002)		(0.014)	(0.024)
Other / Mixed	0.002		-0.011	0.057*
	(0.004)		(0.019)	(0.033)
Admission Index (Standardized)	0.001		-0.003	-0.007
	(0.001)		(0.006)	(0.010)
Financial Aid	-0.001		-0.014	0.015
	(0.002)		(0.013)	(0.022)
High School GPA	-0.003		0.098***	-0.120***
	(0.003)		(0.016)	(0.028)
GPA This Term	0.037***	0.036***		
	(0.001)	(0.005)		
Cumulative GPA	0.028***	0.158***		
	(0.002)	(0.020)		
GPA First Term			0.164***	-0.145***
			(0.006)	(0.014)
STEM Major Ever	0.001		-0.076***	0.294***
	(0.002)		(0.013)	(0.025)
Student Standing Fixed Effects	Y	Y	N	N
Student Fixed Effects	N	Y	N	N
N Observations	51,475	51,475	7,955	5,109
N Students		7,778		

Figure 1: Minimal Correlation with Omitted Predictor Required to Generate Negative Significant Result



Slight positive slopes are due to indirect manipulation of correlation; algorithm may not consider connecting to points directly above and may consider only points above and slightly to the right or left.

Table 6: The Effect of Additional Courses on Performance using Coarsened Exact Matching

Matching at the Student Level using Background				
	(1)	(2)	(3)	(4)
	Standardized Class GPA		Graduation	Time-to-Degree
Classes Attempted	0.004**	0.004**	0.193***	-0.777***
(this term or average)	(0.002)	(0.002)	(0.015)	(0.027)
Control for lag GPA	N	Y	N	N
Control for demographics	N	N	Y	Y
Student Standing Fixed Effects	Y	Y	N	N
Student Fixed Effects	Y	Y	N	N
N Observations	55,178	41,128	7,340	4,750
N Students	7,354	7,062		

Matching at the Student-Term Level using Background and Prior Performance		
	(5)	(6)
	Standardized Class GPA	
Classes Attempted	0.006***	0.003
(this term)	(0.002)	(0.003)
Control for lag GPA	N	Y
Student Standing Fixed Effects	Y	Y
Student Fixed Effects	Y	Y
N Observations	37,115	26,560
N Students	7,058	6,365

Sample sizes differ between (1), (2), (5), and (6) because of the increasing number of lags that must be present to perform matching and estimate the model.