Quality of Students' Course Selections and (Late) College Dropout

Zhengren Zhu*

September 30, 2024

Abstract

I develop a novel measure, *coherence*, for the quality of students' course selections and show that it is closely tied to late college dropout, defined as students dropping out in or after their junior year of college. I leverage the staggered introduction of a web-based advising tool in Texas to estimate the causal effects of the quality of course selections on dropout rates. I find that poor course selections not only increases dropout rate in the current year but also has strong lagging effects on late dropout — a one standard deviation drop in the quality of course selections during the freshmen year not only increases freshmen dropout rate by 7 percentage points, but also increases junior and senior dropout rates by 14 and 10 percentage points. Fortunately, I find that the web-based advising tool studied is highly effective in improving the quality of students' course selections. In addition, I examine factors that affect the quality of students' course selections, and find that parental income, parental education, high school test scores, and ranking of undergraduate programs are positively correlated with higher quality course selections.

1 Introduction

In 2020, the 6-year graduation rate for first-time, full-time, undergraduate students seeking bachelor's degrees at US four-year institutions was 64 percent (NCES, 2022). While the overall graduation rate is indeed low, it hides another disturbing fact — many students drop out of college when they are close to the finish line. Using administrative dataset from the State of Texas, I calculate that while average dropout rates are the highest for freshmen (20%), dropout rates remain significant for later years of enrollment — 10% for sophomores, 8% for juniors, and 11% for seniors. Applying the attrition rates to the nearly 1 million freshmen entering public 4-year universities every year nationwide (National Student Clearing House, 2022), one can estimate that 60,000 juniors, and 70,000 seniors drop out each year. A simple back-of-the-envelope calculation shows that these *late* dropouts are extremely costly

*Department of Economics, Yale University. I would like to thank Joe Altonji, Barbara Biasi, Sarah Cohodes, John Eric Humphries, Brian Jacobs, Costas Meghir, Cormac O'Dea, and Seth Zimmerman for their helpful comments. I would also like to thank seminar participants at the Yale Labor/Public Prospectus Workshops, 2024 AEFP Annual Meeting, and 2024 SOLE Annual Meeting for helpful comments. Greg Branch, Holly Kosiewicz, and Zhixiang Lu provided expert support for the use of the Texas administrative data, for which I am grateful. The administrative data used in this paper is provided jointly by the Texas Higher Education Coordination Board, Texas Education Agency, and Texas Workforce Commission. The results and conclusions of this paper in no ways reflect the official view of the above agencies and the State of Texas. The research would not have been possible without the generous financial support provided by the Cowles Foundation and the Department of Economics at Yale. All remaining errors are mine.

for students as well as for the society. The average annual cost of attendance for a full-time, in-state, student in a US public university is around $27,000 (NCES, 2023), and the median annual earnings for an individual with some college but no degree is around $45,000 (NCES, 2023). Assuming that the cost of attendance reflects the direct cost for students and the public combined, and assuming that students who drop out have the same earnings potential regardless of the number of years enrolled, one can calculate that delaying dropout by one year incurs an additional cost of $72,000 to the society as a whole. As a result, simply shifting the junior-year and senior-year dropouts to dropping out during the sophomore year will save more than $10 billion for the society each year. It is therefore imperative to understand why so many students take so many years to decide to drop out of college.

This paper focuses on the quality of students' course selections as a factor affecting dropout decisions. The motivation for studying the quality of course selections comes from an idea that is almost a priori true: in addition to *how many* courses students take, *what* they take also matters. This is especially true in the US, where college students are given broad flexibility in course selections. However, the importance of students' course selections is understudied because, to the best of my knowledge, there is no quantitative measure that captures the quality of students' course selections. A significant challenge to measuring the quality of students' course selections comes from the fact that a *good* course schedule looks different for students from different majors, different institutions, and different cohorts. I overcome this dimensionality challenge and develop the coherence measure in the following two steps. First, I empirically identify the degree-required courses for each instituion $\times$ major $\times$ cohort cell by examining the transcripts of students in each cell that have successfully attained a degree. Second, I define and calculate the coherence of a student's course selections as the share of degree-required courses that the student has completed in her major. Following this definition, a student gains coherence when she makes course selections that are coherent, in the sense that they are logical and contribute to her degree completion.

Intuition suggests that the quality of course selections might be particularly important in explaining late dropouts. In the US, degree requirements are typically governed by three main factors: accumulating sufficient semester credit hours, maintaining an acceptable GPA, and completing all degree required courses. Unlike semester credit hours and GPA, the quality of students' course selections is not directly reported on their transcripts. In addition, recent survey shows that only around 20-30% of four-year college students were required by their institution to meet with advisors periodically (Inside HigherEd Student Voice Survey, 2023). All of the above make the quality of course selections

difficult to observe, and many students might not realize they have failed to take important courses for their majors until they run degree audits in their senior year or when they start taking advanced courses that have prerequisites. As a result, while low credit accumulation and low GPA in the first years of college would prompt students to dropout immediately, low quality course selections can fly under the radar. Students who are seemingly close to graduating may actually be missing multiple degree-required courses, which prompts late dropouts when detected.

The main empirical goal of this paper is to examine the relationship between late dropout and the quality of course selections, measured by coherence. I leverage recent, staggered, introduction of a web-based advising tool named Degree Works in Texas public universities as a source of exogenous variations in the quality of students' course selections. Degree Works provide automatic degree audit reports when students choose their courses before each semester. Combining event study designs and instrumental variable methods, I find that coherence has a strong lagging effect on students' dropout rates. In particular, a one standard deviation drop in the quality of course selections during the freshmen year not only increases freshmen dropout rate by 7 percentage points, but also increases junior and senior dropout rates by 14 and 10 percentage points. In comparison, low credit accumulation and low GPA are only associated with higher dropout rates in the same year. These results are consistent with low coherence accumulation, or poor course selections, being harder to detect than low GPA and low credits accumulated. Heterogeneity analyses show that low freshmen and sophomore coherence increase late dropout more for STEM majors and first-generation college students. These are in line with STEM majors being "less forgiving" of poor course selections and with students from less privileged backgrounds having a harder time recovering from earlier course selection mistakes. In addition, I perform OLS regressions on community college students, and find similar lagging effects of coherence on dropout rates.

The effects of Degree Works on students' coherence not only serve as the first stage of the main empirical analyses above, but are also of policy interest in and of themselves. Using event study designs, I find that the web-based advising tool significantly improved students' coherence by around 20-50% of a standard deviation. This result is encouraging because it demonstrates that the quality of students course selections can be improved substantially through relatively cheap interventions that do not require significant human capital investment.

In this paper, I also examine factors that contribute to the production of coherence. I find that, unsurprisingly, students with higher high school test scores, parental income, and parental education

make higher quality course selections. Interestingly, I find that students in higher ranking institutions accumulate significantly more coherence each semester, but do not accumulate more credit hours. This pattern holds when controlling for various observable characteristics of students and when using the Dale-Krueger approach to control for endogenous selection into higher ranking institutions (Dale and Krueger, 2002).

This paper contributes to the following three literatures. First, this paper contributes to the literature studying factors explaining students' degree completion. Previous works in this literature have mostly focused on the impact of measures that are relatively easy to quantify, including GPA, credit accumulation, and financial aid available (Dynarski, 2003; Hoffman and Oreopoulos, 2009; Scott-Clayton, 2011a; Fairlie et al, 2014; Stinebrickner and Stinebrickner, 2014; Castleman and Long, 2016). This paper develops a novel measure, coherence, for the quality of students' course selections and provides the first quantitative evidence that the quality of course selections is also an important factor for students' degree completion. In addition, most previous works in this literature have abstracted away from the timing of dropout and do not distinguish between dropouts in the first few years of college and the significantly costlier late dropouts.[1] This paper pays special attention to the timing of dropout and demonstrates how the quality of students' course selections is particularly important for explaining late dropouts.

Second, this paper speaks to the literature on the importance of college advising and student support services. Previous works have found that high-quality student advising and coaching are critical for student success (Bettinger and Baker, 2014; Bahr, 2008; Ehrenberg and Webber, 2010; Scott-Clayton, 2011b; Clotfelter et al, 2018). Scott-Clayton (2011b) describes that "for many students at community colleges, finding a path to degree completion is the equivalent of navigating a shapeless river on a dark night." This paper contributes to this literature in three ways. First, the coherence measure provides a way to quantify the "shape of the river" and can help institutions and policy-makers systematically monitor students' progress towards degree completion. Second, the results on the effect of coherence on dropout reaffirm the importance of advising students to take the right courses, which is arguably the single most important job of academic advisors. Third, the event study analysis on Degree Works provides direct evidence that web-based advising tools can be highly effective in supporting academic advising.

[1] An exception is a recent study by Cohodes and coauthors which reports that even highly qualified college students in a highly selective institution drop out after more than three years in college (Cohodes, Ho, and Robles, 2022) at a non-trivial frequency.

Third, this paper is related to two literatures that study students' fields of study. First, this paper is related to the vibrant literature that studies the returns to different college majors (Hastings et al, 2013; Kirkeboen et al, 2016; Bleemer and Mehta, 2022; Andrews et al, 2024). Since the coherence measure is major-specific, a student's coherence in a major can be taken as a one-dimensional summary of the student's course-taking towards a degree in the major. A student's coherence in a major can therefore be used as a proxy for her preference and ability in that major, and can be used as a powerful control variable when studying the impact of fields of study on student outcomes. Second, this paper is also related to a number of studies that discuss the college-major versus the college-then-major design of the higher education system. The central question of this literature is whether students may benefit from choosing their fields of study before or after enrolling in college. Previous works in this literature have mostly focused on the trade-off between skill specialization and skill mismatches (Arcidiacono, 2004; Stinebrickner and Stinebrickner, 2013; Borden and Fu, 2015). This paper contributes to this debate by showing that students who choose their degree plans later might be at a higher risk of not finishing their degrees due to incoherent course selections during their first years in college. As a result, a college-then-major design should come with guardrails for students to take sensible courses while exploring different majors.

The remainder of the paper proceeds as follows: Section 2 introduces the Texas administrative data used in this paper, defines the coherence measure, and explains how the coherence measure is calculated using the Texas administrative data. Section 3 provides institutional backgrounds related to the web-based advising tool studied, Degree Works, and describes the main empirical strategies. Section 4 presents the main empirical results on the effects of the quality of course selections on dropout rates. Section 5 presents the results on factors affecting the quality of students' course selections. Finally, Section 6 concludes and provides directions for future research.

2 Data and the Construction of the Coherence Measure

2.1 The Texas Administrative Data

This paper uses a large administrative data from Texas, which links post-secondary education data from the Texas Higher Education Coordination Board (THECB), to K-12 data from the Texas Education Agency (TEA), and to labor-market data from the Texas Workforce Commission (TWC). This

allows researchers to observe an individual's progression from high school, to college, and to the labor market, so long as the individual remains in Texas.

The TEA data provides enrollment records, demographic backgrounds, course completion, graduation records, and attendance records for all Texas public high school students since 1993. The TEA data also includes students' performance in the state-wide standardized exams required for high school graduation. The THECB data provides information on all college students enrolled in Texas public higher education institutions since 1992.[2] The Texas public higher education system is composed of 38 four-year colleges and 50 public community colleges, and is the second largest in the US (after California) in terms of number of enrollment. The THECB data contains records of students' demographic background, enrollment, course selections and course grades, degree attainment, as well as financial aid and student loan. The availability of transcript level data started in the 2011-2012 academic year. Since the calculation of students' coherence requires their transcript information, I restrict my main analysis to students who started college in or after the 2011-2012 academic year. Since the calculation of students' coherence also requires a significant number of students from their entering cohort to have attained a degree, I further restrict my main analysis to students who started college in or before the 2018-2019 academic year.

2.2 Coherence Measure: Definition and Measurement

I develop the coherence measure to quantify the quality of students' course selections. Intuitively, a high-quality course selection is one that is logical and contributes to a student's degree completion. Following this intuition, the coherence measure is designed to capture a student's progress towards completing degree-required coursewoks and is defined as the share of required courses that the student has completed. Formally,

$$c_{it}^{j} = \begin{cases} \frac{h_{it}^{j}}{n^{j}} & \text{if } h_{it}^{j} < n^{j} \\ 1 & \text{if } h_{it}^{j} \geq n^{j} \end{cases} \tag{1}$$

[2] While the data has information on students from both public and private institutions, I restrict my analysis to students in public institutions since critical information including transcript records are not available for private colleges.

where c_{it}^j denotes the coherence of student i at time t in major j, h_{it}^j denotes the number of credits in required courses for major j that student i has completed by time t, and n^j denotes the total number of credits in required courses for major j. By definition, the coherence measure is a measure between 0 and 1, where having a coherence of 1 simply means the student has completed all required courseworks in major j.

The overall coherence of student i is then defined as her coherence in her declared major j^* if j^* is declared. If the student has not yet declared a major, her coherence is defined as her highest coherence amongst all majors:

$$c_{it} = \begin{cases} c_{it}^{j^*} & \text{if } j^* \text{ declared} \\ \max_{j \in J} \ c_{it}^j & \text{if } j^* \text{ not declared} \end{cases} \tag{2}$$

The observation of students' transcript level data makes measuring coherence possible. However, a significant challenge in measuring coherence remains: the more than 80 colleges and dozens of college majors create thousands of institution $\times$ major pairs that each has a distinct set of required courses. In addition, degree requirements and course labeling have potentially changed over time. Manually coding the degree requirements for students is therefore close to impossible and prone to severe human error. To solve this challenge, I empirically identify the major requirements using the following procedure. First, I identify the population of students who attain degree J from institution M in cohort T, $S_{J,M,T}$. Second, I identify the courses that more than 90% of $S_{J,M,T}$ took, and label these courses as the set of required courses, $C_{J,M,T}^R$. Third, I identify courses that more than 30% of $S_{J,M,T}$ took, and label these courses as the set of elective courses, $C_{J,M,T}^E$. Fourth, I count the number of elective courses that students in $S_{J,M,T}$ took and take the 30th percentile value of that distribution, $n_{J,M,T}^E$. The tuple $\{C_{J,M,T}^R, C_{J,M,T}^E, n_{J,M,T}^E\}$ then characterizes the degree requirement of degree J from school M in cohort T. In particular, students need to complete all courses in the set $C_{J,M,T}^R$, and at least $n_{J,M,T}^E$ courses in the set $C_{J,M,T}^E$.

I present evidence that this empirical procedure captures students' degree requirement with significant accuracy. To examine the accuracy of the empirical procedure for identifying degree requirements, I compare the required courses identified by the procedure and the actual required courses listed on institutions' websites for a random set of institution $\times$ major pairs.[3] Table 1 shows the accuracy

[3]The 2019 entering cohort was used for this procedure.

rates of 16 institution × major pairs I tested. The audits of all pairs show very high accuracy of the empirically identified degree requirements. The only institution × major pair with slightly lower accuracy is UT Austin x Economics (74%). The reason behind the relatively low accuracy is because many UT Austin economics majors enroll in business-related courses that are not necessarily required by the economics major.

In addition, I calculate a separate coherence measure for completion of the Texas Core Curriculum (TCC). The TCC is a 42 semester credit hour core curriculum for all undergraduate students in Texas public higher education institutions (both community colleges and four-year colleges). The TCC is organized in nine component areas: communication; mathematics; life and physical science; language, philosophy and culture; creative arts; American history; government and political science; social and behavioral sciences; and an optional area. The TCC requirement is mandated by state law to be transferrable between institutions. The TCC required courses are taken directly from THECB official websites.

3 Institutional Backgrounds and Empirical Strategy

3.1 Degree Works

Degree Works, hereafter DW, is an online degree audit tool that allows students and their academic advisors to monitor students' progress towards degree completion and plan future academic courseworks. DW has two main functionalities. The first function is degree audit — DW automatically matches a student's transcript with the student's degree plan and calculates the student's progress towards degree completion. Figure 1 is a snapshot of the DW tool for Texas State University, which shows that the student has completed 89% of the degree requirements for a BS in Clinical Lab Science degree. The second function is "what-if" analyses — DW allows students to input hypothetical course selections for future semesters and calculate the resulting progress towards degree completion. While institutions highlight that DW is not a substitute for academic advisors, it does replicate many of their roles and allows students to be more informed when choosing courses. Moreover, DW, like other web-based tools, are easier to access than academic advisors since it does not require making an appointment. As a result, it is reasonable to believe that DW improves the quality of course selections and the coherence of students.

DW is available for more than one third of the 38 public four-year institutions in Texas. Table 2

provides a list of Texas public 4-year institutions that have DW implemented, and the years in which they implemented DW. I leverage the staggered implementation of DW as a source of exogenous variations in students' coherence to identify the causal effects of coherence on dropouts. The main variations for identification will come from the five institutions that implemented DW during the main sample observation window of 2012-2019. These institutions are UT San Antonio, Tarleton State University, Angelo State University, Midwestern State University, and Prairie View A&M University. Combined, these five institutions enroll around 10% of all public university students in Texas.

3.2 Empirical Strategy — Effects of Coherence on Dropout

The first focus of the empirical analysis is to study how coherence accumulation impacts students' dropout rates. For student i enrolled in institution j, our main outcome of interest is DropOut_{it}, which is an indicator variable for whether the student dropped out during year t. The baseline OLS regression is:

$$\text{DropOut}_{it} = \beta_0 + \beta_{coh}\text{coherence}_{it} + \beta_{sch}\text{credit}_{it} + \beta_{gpa}\text{gpa}_{it} + \beta_x X_{it} + \beta_s S_j + \beta_m M_{it} + \epsilon_{it} \quad (3)$$

where coherence_{it},credit_{it}, and gpa_{it} are the coherence gained, number of credits accumulated, and gpa for student i during year t. The regression controls for a vector of student observables X_{it}, which includes student's high school standardized test score, gender, ethnicity, parental income, parental education, and part-time enrollment status. It also controls for student's school of enrollment S_j, and major enrolled, M_{it}. For the analyses, I restrict to students who are not transfer students from other two-year or four-year colleges.

In addition to the contemporaneous effects of coherence on dropout rates, I am also interested in examining whether coherence gains in the earlier years of college enrollment might have lagging effects on dropout rates. While credits accumulated and GPA are clearly displayed on transcripts after each semester, coherence is typically not displayed and harder to observe for students. Students may find out about their freshmen coherence accumulation when they try to enroll in an advanced level course that requires multiple prerequisite, when they meet with an academic advisor in later semesters, or when they apply for a degree in their senior year. If this is the case, earlier coherence

accumulation could fly under the radar and affect later years' dropout decisions. For this purpose, I extend equation (3) to include lags of coherence, semester credit hours, and GPA. Denote a student's academic standing as $Y \in \{1, 2, 3, 4\}$, where 1, 2, 3, and 4 represent freshmen, sophomore, junior, and senior, respectively. Then:

$$\text{DropOut}_i^Y = \beta_0 + \sum_{t=1}^{Y} \beta_{coh}^{tY} \text{coherence}_i^t + \sum_{t=1}^{Y} \beta_{sch}^{tY} \text{credit}_i^t + \sum_{t=1}^{Y} \beta_{gpa}^{tY} \text{gpa}_i^t + \beta_x X_i + \beta_s S_j + \beta_m M_i + \epsilon_i \quad (4)$$

where coherence_i^t,credit_i^t, and gpa_i^t are the coherence gained, number of credits accumulated, and gpa for student i during enrollment year t. β_{coh}^{tY}, β_{sch}^{tY}, and β_{gpa}^{tY} are the effects of coherence gained, number of credits accumulated, and GPA in year-t on the dropout rates in year-Y.

3.2.1 Event Study Design and Instrumental Variables Method

Despite the rich set of control variables available, it is likely that coherence accumulation is correlated with unobservable student characteristics that contribute to degree completion. For example, students who excel in planning ahead of time are more likely to make more coherent course selections and are also less likely to drop out. To overcome this endogeneity issue, I leverage the staggered introduction of DW, which is the online degree audit tool introduced in the previous subsection. A parallel trends assumption is needed — I assume that the trends of student outcomes for the universities that introduced DW would have been the same as those that did not if DW was not introduced. Figure 2 compares the pre-DW trends of dropout rates at the five treated institutions and the trends of dropout rates at institutions that do not have DW (never-treated group). The figures suggest that there is no discernable patterns in the pre-trend comparisons, indicating that the parallel trends assumption is acceptable. I follow Borusyak et al (2021), and estimate the treatment effects of DW on students' coherence accumulation using a TWFE regression that includes group and time fixed effects, as well as fixed effects for all treated institution $\times$ year cells. Under the parallel trends assumption, this TWFE approach allows for dynamic treatment effects under staggered treatments, and identifies the institution $\times$ year — specific treatment effects of DW. The TWFE regression is as follows:

$$\text{coherence}_{it} = \gamma_0 + \gamma_1 X_{it} + \gamma_2 M_{it} + \lambda_j + \delta_t + \Gamma \text{D}_{jt} + \nu_{it} \quad (5)$$

where λ_j is the institution FE and δ_t is the year FE. D_{jt} is a set of FE for each $\{j,t\}$ cells for which DW is available. Under parallel trends, the vector of coefficients Γ provides unbiased estimates for the treatment effects of DW for students in the corresponding $\{j,t\}$ cell.

I also estimate a TWFE specification that separately estimates treatment effects for freshmen, sophomores, juniors, and seniors. This specification leverages the fact that students at the same institution received different levels of exposure to DW depending on their cohort. For example, Angelo State University introduced DW in 2016. As a result, for a student who finishes college in four years, she would have been exposed to DW for her entire college years if she entered in 2016, but would only have been exposed for her junior and senior years if she entered in 2014. Following this intuition, let c denote student i's entering cohort. For $Y \in \{1,2,3,4\}$, I run regressions as follows:

$$\text{coherence}_i^Y = \gamma_0^Y + \gamma_1^Y X_i + \gamma_2^Y M_i + \lambda_j + \delta_c + \Gamma^Y D_{jc}^Y + \nu_{ic} \tag{6}$$

Compared to equation (5), the difference in this specification is that δ_c is a cohort FE instead of a year FE, and D_{jc}^Y is a set of FE for each $\{j,c\}$ cells for which DW is available in the cohort's year Y of enrollment. Continuing on the previous example, the cell $\{\text{Angelo State University}, 2016\}$ would be included in D_{jc}^1, D_{jc}^2, D_{jc}^3, and D_{jc}^4, while $\{\text{Angelo State University}, 2014\}$ would only be included in D_{jc}^3 and D_{jc}^4.

The TWFE specifications in equations (5) and (6) will be used as first stages for IV estimations for equations (3) and (4), with the sets of FE, D_{jt} and D_{jc}^Y, used as excluded instruments. It is important to note that, aside from serving as a first stage to estimating the effect of coherence on dropout, the TWFE regressions are also of policy interest in and of itself. The coefficient estimates will provide insights on whether simple web-based advising tools such as DW can be effective in improving the quality of students' course selections.

3.3 Empirical Strategy — Production of Coherence

Another goal of the paper is to understand factors that contribute to the production of coherence. The baseline regression is:

$$\text{Coherence}_{it} = \alpha_0 + \alpha_x X_{it} + \alpha_s S_j + \alpha_m M_{it} + \epsilon_{it} \tag{7}$$

where X_{it} include gender, ethnicity, parental income, parental education, and high school test scores. Similar to previous specifications, S_j are fixed effects for institutions enrolled, and M_{it} are fixed effects for the major a student is enrolled in. To compare coherence production to the accumulation of semester credit hours, I also run regressions following equation (5), but with the dependent variable substituted with credits accumulated.

To examine whether quality of institution is related to coherence accumulation, I proxy quality with the US News and World Reports ranking of institutions, rank_j, and estimate:

$$\text{Coherence}_{it} = \alpha_0 + \alpha_x X_{it} + \alpha_r \text{rank}_j + \alpha_m M_{it} + \Omega_i + \epsilon_{it} \tag{8}$$

The ranking of institution is most likely correlated with unobservable student characteristics. To control for selection issues, I follow Dale and Krueger's (2002) idea of addressing selection into specific institutions by controlling for the programs a student applied to and was admitted to. Ω_i includes a vector of indicators that specifies the institutions that student i applied to and were admitted to. Given the availability of data, Ω_i only considers students' application portfolio among Texas public four-year institutions.[4]

4 Results

Before turning to the main results, I first illustrate the importance of the coherence measure by comparing the accuracy of predicting the graduation of a four-year college student when using semester credit hours accumulated, GPA, coherence, and various combinations of the three measures.

The results are presented in Table 3.1. I find that while students' credits accumulated and GPA strongly predict four-year college graduation, the prediction is not perfect — 79% of students who have at least 120 semester credit hours and a GPA higher than 2.0 graduates within a year. Once coherence is also taken into account, however, the prediction improves significantly — more than 90% of students who have at least 120 semester credit hours, a GPA higher than 2.0, and coherence of 1 receives a bachelor's degree within a year. In other words, the coherence measure improves the predictive power of student progress towards bachelor's degree completion by around 15%, compared

[4] Mountjoy and Hickman (2021) also use the Dale-Krueger approach on the Texas adminsitrative data.

to when only credits accumulated and GPA were used.

Repeating the same exercise on community college students (reported in Table 3.2), I find that taking into account coherence also improves predictive power of associate's degree completion by around 15%. The overall predictive power of this exercise is weaker than in four-year colleges, because many students choose to transfer to a four-year college before attaining an associate's degree.

4.1 Event Study Results: Effectiveness of DW on Improving Quality of Course Selections

The staggered introductions of DW in Texas institutions are used as exogenous shocks to students' coherence. This not only facilitates the identification of the causal effects of coherence on dropout rates, but also has policy interest in and of itself. Figure 3 presents the institution $\times$ year $-$ specific treatment effects of DW on students' coherence, following the event study design in equation (5). The figure shows that DW significantly improved students' coherence. 27 out of the 33 institution $\times$ year cells had statistically significant positive treatment effects, and only one cell (Angelo State University in the first year after implementation) had a (marginally) statistically significant negative treatment effects. Additionally, there seems to be an increase in the treatment effects in the years after implementation of DW. Across the five institutions, the treatment effects peak around 0.02 on the low end and 0.05 on the high end. Given that the standard deviation of coherence accumulation per semester is around 0.1, the positive effects of DW on coherence reaches 20% to 50% of a standard deviation once fully implemented.

The institution $\times$ year $\times$ academic standing $-$ specific treatment effects of DW are estimated following equation (6) and the results are presented in Figure 4. While the error bars are not reported for the clarity of the figures, 11 of the 18 treatment cells for freshmen coherence had statistically significant positive treatment effects. 12 of the 23 treatment cells for sophomore coherence had statistically significant positive treatment effects. 16 of the 28 treatment cells for junior coherence had statistically significant treatment effects. 16 of the 25 treatment cells for senior coherence had statistically significant positive treatment effects. Across the board, only 4 of the 94 treatment cells had statistically significant negative treatment effects. Similar to the pattern found in the institution $\times$ year $-$ specific treatment effects, the treatment effects of DW appear to increase over time.

I conclude from these results that (a) web-based advising tools such as DW can be highly effective in improving the quality of students' course selections; and (b) the staggered introduction of DW is a

promising instrument for studying the causal effects of coherence on dropout rates.

4.2 Effect of Coherence on Dropout Rates

I now turn to the results on the effects of coherence on dropout rates. First, I present the results of regressions in the form of equation (3), which examine the effects of current period coherence, semester credit hours, and GPA on persistence. Table 4 presents the results separately for students in their freshmen, sophomore, junior, and senior years. Since coherence, semester credit hours, and GPA have different scales, the coefficients were translated to represent the impact of a one standard deviation increase in each of the three factors. One can see that increases in all three factors are correlated with a decrease in the likelihood of dropping out. Although the absolute value of the coefficients for coherence appears smallest among the factors, the effects are still significant. A one standard deviation increase in coherence is associated with a decrease in the likelihood of dropout by around 0.7 percentage points for sophomores and seniors, and around 2 percentage points for freshmen and juniors. This represents about a 10-20% decrease in dropout at each year of enrollment.

As discussed in Section 3.2, intuition suggests that earlier coherence accumulation could have lagging effects on later years' dropout rates. To test this hypothesis, I run regressions following specification of equation (4). The results are presented in Table 5 and confirm the earlier hypothesis that coherence accumulation has a significant lagging effect on dropout. Conditional on reaching junior year or senior year status, the freshmen and sophomore coherence accumulation have as much, if not stronger, correlation with dropout decisions as the junior and senior coherence accumulations. In contrast, the credit accumulation and GPA of freshmen and sophomore year have no meaningful lagging effects on later drop out decisions.

To examine if coherence's lagging effects on dropout rates are causal, I run instrumental variable regressions that use the staggered introductions of DW as instruments for coherence following the discussion in Section 3.2.1. The results are presented in Table 6, and show patterns that are consistent with the OLS results. The first row of the Table shows that freshmen coherence significantly reduces dropout rates in junior and senior years. The point estimates are three to fives times larger than the OLS estimates. The estimates of sophomore coherence on junior and senior year dropouts are in-line with the OLS estimates, but only marginally statistically significant given the larger standard error. It is not surprising that the absolute value of the causal estimates are larger than those of the OLS

estimates. The IV estimates identify the effects of coherence on dropout for the complier group, who are those that make better course selections in the presence of DW. It is likely that these are the students that need academic advising the most and are the most at risk of dropping out.

In addition to the average causal effects of coherence on dropouts, two heterogeneity results are reported in Table 7 and are worth discussing. First, I find that the lagging effects of coherence on late dropouts are more significant for STEM majors. This is consistent with the fact that STEM majors, especially engineering majors, have highly structured degree plans with more required courses and more courses with specific prerequisites. As a result, these majors are potentially "less forgiving" when students make poor decisions on course selections early on in college. Second, I find that the lagging effects of coherence are generally more significant for first generation college students. This heterogeneity result suggests that students with less resources and informal advising from college-educated parents have a harder time recovering from poor course selections earlier on in college.

The significant lagging effects of coherence on dropout implies that improving coherence can have a significant impact on student persistence especially during their later years in college. A related policy question of interest is how much a simple web-based intervention like DW improve students persistence? I perform a simple simulation exercise to answer this question. First, I calcuate an average "steady-state" treatment effect of DW on students' coherence. For each of the five treated schools, I take the treatment effect of DW on the last observed treated cohorts for freshmen, sophomore, junior, and senior coherence, and take these as the "steady-state" treatment effect for each of the institution. I then compute the average treatment effect by weighting with each school's enrollmnet. The average "steady-state" treatment effects of DW are 0.033 for freshmen, 0.046 for sophomores, 0.032 for juniors, and 0.030 for seniors. Second, I calculate the counterfactual dropout rates for freshmen, sophomores, juniors, and seniors by multiplying the steady-state treatment effects with the coefficient estimates in Table 6. Third, I calculate the share of students who persist after each year of enrollment with the baseline and counterfactual dropout rates. The result is presented in Figure 5. The result shows that DW can increase graduation rate by close to 20% from 59% to 70%. Moreover, most of the improvements come from reducing the most costly types of dropouts during the junior and senior years.

4.3 Effect of Coherence on Community College Student Outcomes

I repeat the regressions from equations (3) and (4) on community college students, and the results are presented in Table 8. Given community colleges' dual purpose of providing associate degrees and facilitating transfer to four-year colleges, I define dropout as leaving college without either transferring or an associate degree. The results in columns (1) - (2) suggest that coherence decreases the likelihood of dropping out, although the effects are smaller than those of semester credit hours and GPA. Columns (3) - (8) break out the effects of coherence, credits, and GPA into effects on associate degree attainment, transfer to four-year colleges, and bachelor's degree attainment. The results suggest that, while coherence accumulation does not increase the likelihood of associate degree attainment, it significantly increases the likelihood to transfer to a four-year college and the likelihood of attaining a bachelor's degree. This partly reflects the fact that, among community college students who transfer to four-year colleges, 57% transfer without attaining an associate degree.

Given the fact that many community college students are interested in transferring instead of associate degrees, I repeat the above analyses when substituting the coherence measure with the students' coherence for the TCC. Because TCC courses are guaranteed to be transferrable between community colleges and four-year college, the coherence measure for TCC may be more appropriate for community college students' that are interested in transferring. The results, not reported, are not qualitatively different from the results in Table 8. Overall, the results suggest that while taking the right courses in community college may not be a challenge for the purpose of attaining an associate degree, it is important to the likelihood of transferring to a four-year college and attaining a bachelor's degree. It should be noted, however, that the results on community college students are correlational and may be subject to endogeneity issues.

5 Production of Coherence

In this section, I turn to the results on factors that contribute to the production of coherence. First, I test for impact of individual-level factors, including parental income, first-generation college student status, and high school test scores. Second, I examine whether quality of institution affect students' coherence accumulation.

To examine the impact of individual-level factors, I run regressions following equation (7). The estimates for freshmen, sophomores, juniors, and seniors are reported separately in Table 9. Students

with higher high-school test scores accumulate coherence faster — a ten percentile increase in HS test score is associated with a 0.4 pp, 0.3 pp, 0.3 pp, and 0.1 pp improvements in coherence accumulation each semester in freshmen, sophomore, junior, and senior years, respectively. Students with higher parental income also accumulate coherence faster — the effect of a three decile increase in parental income is roughly equivalent to that of a one decile increase in HS test score. First generation college students accumulate less coherence and the gap to students with college-educated parents is roughly equivalent to the impact of a 5 percentile increase in HS test scores. This result suggests that parental resources may be a substitute to academic advising when it comes to the quality of course selections.

I run regressions that control for students' application / admission portfolios following equation (8). The results on the effect of four-year college ranking on coherence and credit accumulation are reported in Table 10. The results show that higher ranked institutions are better at helping students accumulate coherence faster. An increase in ranking by 100 (i.e., moving from a top 100 college to a top 200 college) reduces coherence accumulation by 0.4 pp, 0.2 pp, 0.1 pp, and 0.1 pp for freshmen, sophomores, juniors, and seniors, respectively. Comparing to results in Table 9, this is similar to the effects of a one decile increase in high school test score or a 3 decile increase in parental income.

Although already mentioned in Section 4.3, it is worth mentioning here that the event study results on the impact of DW on coherence suggest that improved academic advising is highly effective in improving students' coherence.

6 Conclusions and Directions for Future Research

In this paper, I formalize the widely accepted idea that, in addition to the number of courses taken, students' course selections also matters. I do so by defining and constructing a novel measure, coherence, for the quality of students' course selections. Leveraging the rich Texas administrative data, I calculate the coherence measure for all Texas college students and find that the coherence measure significantly improves the prediction for degree completion compared to when only GPA and credits accumulated are used. Leveraging the staggered implementation of DW, I find evidence that coherence has strong lagging effects on dropout rates and can help explain the costly but prevalent late dropout phenomenon. Slow coherence accumulation in earlier years of college significantly increases the likelihood of dropout in junior and senior years. This lagging effect is not present for GPA and credit accumulation. Unlike failing courses and receiving a low GPA, taking the wrong courses is

harder to detect. As a result, many students may not realize they have low coherence accumulation until they take advanced courses that requires prerequisites or when they go through degree audits, leading to late dropouts.

Examining factors contributing to the quality of students' course selections, I find that high school test scores, parental income, and parental education are significantly correlated with coherence accumulation, implying that cognitive skills and parental resources help students choose the right courses. I find evidence that institutions also have a significant role to play in the production of coherence — students enrolled in higher ranking institutions tend to accumulate coherence faster. The event study results on the implementation of DW show that DW significantly improved students' coherence, implying that simple web-based interventions that target academic advising can be instrumental for keeping students on track with their course selections.

As the first paper to consider the measurement of the quality of students' course selections, this paper leaves open several areas for future research. First, the methodology of computing coherence of students' course selections can be extended to high school. This will allow researchers and practitioners to move beyond standardized test scores and GPA to identify priority high school courses that are particularly important for students' educational and labor-market outcomes. Second, this paper leaves the issue of the sequence of course-taking open. It is well known among academic advisors that some courses, such as calculus and introduction to microeconomics, should be completed early in college because they serve as prerequisites for required courses in many different majors. Identifying these courses and quantifying the importance of correct sequencing of courses is an important topic for future study. One potential way is to compute an academic-standing-specific coherence by running this paper's proposed procedure separately for freshmen, sophomores, juniors, and seniors. For example, one can identify the common courses that all successful BA in Economics students in UT Austin took during their freshmen year, and calculate freshmen coherence for Economics students at UT Austin accordingly. Third, since the coherence measure is major-specific and reflect students' ability and preference in a specific major, it has the potential of being a powerful control variable when studying the effects of college majors.

References

Andrews, Rodney J., Scott A. Imberman, Michael F. Lovenheim, and Kevin Stange. 2024. "The Returns to College Major Choice: Average and Distributional Effects, Career Trajectories, and Earnings Variability." *Review of Economics and Statistics*, forthcoming.

Arcidiacono, Peter. 2004. "Ability Sorting and the Returns to College Major." *Journal of Econometrics* 121 (1-2): 343-375.

Bahr, Peter Riley. 2008. "Cooling Out in the Community College: What is the Effect of Academic Advising on Students' Chances of Success?" *Research in Higher Education* 49: 704-732.

Bettinger, Eric and Rachel Baker. 2014. "The Effects of Student Coaching: An Evaluation of a Randomized Experiment in Student Advising." *Educational Evaluation and Policy Analysis* 36 (1): 3-19.

Bleemer, Zachary, and Aashish Mehta. 2022. "Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major." *American Economic Journal: Applied Economics* 14 (2): 1-22.

Bordon, Paola, and Chao Fu. 2015. "College-Major Choice to College-Then-Major Choice." *Review of Economic Studies* 82 (4): 1247-1288.

Borusyak, Kirill, Xaviel Jaravel, and Jann Spiess. 2024. "Revisiting Event-Study Designs: Robust and Efficient Estimation." *Review of Economic Studies*, forthcoming.

Castleman, Ben L., and Bridget Terry Long, 2016. "Looking Beyond Enrollment: The Causal Effect of Need-Based Grants on College Access, Persistence, and Graduation." *Journal of Labor Economics* 34 (4): 1023-1073.

de Chaisemartin, Clement, and Xavier D'Haultfoeuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110 (9): 2964-2996.

de Chaisemartin, Clement, and Xavier D'Haultfoeuille. 2022. "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey." NBER Working Paper: 29691.

Clotfelter, Charles T., Steven W. Hemelt, and Helen F. Ladd. 2018. "Multifaceted Aid for Low-Income Students and College Outcomes: Evidence from North Carolina." *Economic Inquiry* 56 (1): 278-303.

Cohodes, Sarah R., Helen Ho, and Silvia C. Robles. 2022. "STEM Summer Programs for Underrepresented Youth Increase STEM Degrees." NBER Working Paper: 30227.

Dale, Stacy Berg, and Alan B. Krueger. 2002. "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *The Quarterly Journal of Economics* 117(4): 1491-1527.

Dynarski, Susan M. 2003. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion." *American Economic Review* 93 (1): 279-288.

Ehrenberg, Ronald, and Douglas Webber. "Do Expenditures Other than Instructional Expenditures Affect Graduation and Persistence Rates in American Higher Education." *Economics of Education Review* 29 (6): 947-958.

Fairlie, Robert W., Florian Hoffman, and Philip Oreopoulos. 2014. "A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom." *American Economic Review* 104 (8): 2567-91.

Hastings, Justine S., Christopher A. Neilson, and Seth D. Zimmerman. 2013. "Are Some Degrees Worth More than Others? Evidence from College Admission Cutoffs in Chile." NBER Working Paper:

19241.

Hoffman, Florian, and Philip Oreopoulos. 2009. "A Professor Like Me: The Influence of Instructor Gender on College Achievement." *Journal of Human Resources* 44 (2): 479-494.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. 2016. "Field of Study, Earnings, and Self-Selection." *Quarterly Journal of Economics* 131 (3): 1057-1111.

Mountjoy, Jack, and Brent R. Hickman. 2021. "The Returns to College(s): Relative Value-Added and Match Effects in Higher Education. NBER Working Paper: 29276.

Scott-Clayton, Judith. 2011a. "On Money and Motivation: A Quasi-Experimental Analysis of Financial Incentives for College Achievement." *Journal of Human Resources* 46 (3): 616-646.

Scott-Clayton, Judith. 2011b. "The Shapeless River: Does a Lack of Structure Inhibit Students' Progress at Community Colleges?" CCRC Working Paper No.25. New York: Columbia University.

Stinebrickner, Todd, and Ralph Stinebrickner. 2012. "Learning about Academic Ability and the College Dropout Decision." *Journal of Labor Economics* 30 (4): 707-748.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2013. "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout." *Review of Economic Studies* 81 (1): 426-472.

Appendix A: Tables and Figures

Table 1: Accuracy of Audit of 16 Institution x Degree Pairs

Institution x Degree	**% in Catalog**
UT Austin x Business	88%
UT Austin x Biology	94%
UT Austin x Economics	74%
UT Austin x Communications	95%
U North Texas x Computer Science	100%
U North Texas x Biology	100%
UT Arlington x Business	88%
UT Arlington x Psychology	93%
Sam Houston State x Criminal Justice	98%
U Houston x Psychology	88%
UT Rio Grande Valley x Nursing	98%
UT San Antonio x Business	91%
UT El Paso x Mechanical Engineering	96%
Texas Tech x Biology	85%
UT Dallas x Computer Science	86%

Notes: This table presents the accuracy of the empirically identified degree requirements for 16 random institution x degree pairs. The audit focus on the degree requirements for the 2023 graduation cohort since I compare the empirically identified requirements to the current degree plans posted on the institutions' websites. The numbers presented here calculates the share of the empirically identified degree required courses that are found in the actual degree plans.

Table 2: Institutions with DegreeWorks and Year of Implementation

Institution	Year Implemented
Texas A&M - Commerce	2010
Texas A&M - Corpus Christi	2010
Sam Houston University	2011
Texas A&M - Kingsville	2011
Texas State University	2011
Texas A&M - Texarkana	2012
Texas A&M - San Antonio	2012
Texas Tech University	2012
UT - Rio Grande Valley	2012
UT - San Antonio	2014
Tarleton State University	2015
Angelo State University	2016
Midwestern State University	2016
Prairie View A&M University	2016

Table 3.1: Predictive power of coherence on graduation — 4-yr colleges

	Probability of Attaining Bachelor's Degree This Year			
	74%	20%	79%	91%
Semester Credit Hours $\geq$ 120	✓		✓	✓
GPA $\geq$ 2.0		✓	✓	✓
Coherence = 1				✓

Notes: This table presents the predictive power of different factors for four-year college graduation. The first column presents the share of students with more than or equal to 120 semester credit hours who graduates within a year of the observed semester. The second column presents the share of students with a GPA of at least 2.0 that graduates within a year of the observed semester. The third column presents the share of students with more than or equal to 120 semester credit hours and a GPA of at least 2.0 that graduates within a year of the observed semester. The fourth column presents the share of students with more than or equal to 120 semester credit hours, a GPA of at least 2.0, and a coherence of 1 that graduate within a year of the observed semester.

Table 3.2: Predictive power of coherence on graduation — community colleges

	Probability of Attaining Bachelor's Degree This Year			
	41%	13%	53%	61%
Semester Credit Hours $\geq$ 60	✓		✓	✓
GPA $\geq$ 2.0		✓	✓	✓
Coherence = 1				✓

Notes: This table presents the predictive power of different factors for community college graduation. The first column presents the share of students with more than or equal to 60 semester credit hours who graduates within a year of the observed semester. The second column presents the share of students with a GPA of at least 2.0 that graduates within a year of the observed semester. The third column presents the share of students with more than or equal to 60 semester credit hours and a GPA of at least 2.0 that graduates within a year of the observed semester. The fourth column presents the share of students with more than or equal to 60 semester credit hours, a GPA of at least 2.0, and a coherence of 1 that graduate within a year of the observed semester.

Table 4: Effects of Current-Year Coherence, Credits, and GPA on Dropout

	Dropout			
	Freshmen	Sophomore	Junior	Senior
	(1)	(2)	(3)	(4)
Coherence	-0.022	-0.007	-0.017	-0.007
	(0.001)	(0.001)	(0.001)	(0.001)
Semester Credit Hours	-0.052	-0.052	-0.045	-0.012
	(0.003)	(0.003)	(0.000)	(0.003)
GPA	-0.281	-0.119	-0.087	-0.102
	(0.001)	(0.001)	(0.001)	(0.001)
N	186,445	144,906	136,889	157,275

Notes: This table presents the results of regressions in the form of equation (3). Columns (1) — (4) presents result for freshmen, sophomore, junior, and senior years respectively. Each regression restricts to the sample of students who reach the corresponding year. The dependent variable is an indicator for whether the student drops out during that year. The main explanatory variables are the coherence accumualted, semester credit hours accumulated, and GPA per semester in the corresponding year. The main explanatory variables are scaled to their standard deviations in each year. Robust standard errors are reported in parentheses.

Table 5: Effects of Lagging and Current Coherence, Credits, and GPA on Dropout

	Dropout			
	Freshmen	Sophomore	Junior	Senior
	(1)	(2)	(3)	(4)
Coherence - Freshmen	-0.022	-0.021	-0.028	-0.036
	(0.001)	(0.001)	(0.001)	(0.001)
Coherence - Sophomore		-0.013	-0.019	-0.029
		(0.001)	(0.001)	(0.001)
Coherence - Junior			-0.027	-0.027
			(0.001)	(0.001)
Coherence - Senior				-0.023
				(0.001)
Semester Credit Hours - Freshmen	-0.052	0.003	-0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.003)
Semester Credit Hours - Sophomore		-0.052	-0.006	-0.006
		(0.003)	(0.003)	(0.003)
Semester Credit Hours - Junior			-0.042	-0.016
			(0.002)	(0.000)
Semester Credit Hours - Senior				-0.003
				(0.000)
GPA - Freshmen	-0.281	-0.021	0.000	0.000
	(0.001)	(0.001)	(0.002)	(0.002)
GPA - Sophomore		-0.123	0.003	-0.008
		(0.002)	(0.002)	(0.002)
GPA - Junior			-0.084	-0.016
			(0.002)	(0.002)
GPA - Senior				-0.084
				(0.002)
N	186,445	144,906	136,889	157,275

Notes: This table presents the results of regressions in the form of equation (4). Columns (1) — (4) presents result for freshmen, sophomore, junior, and senior years respectively. Each regression restricts to the sample of students who reach the corresponding year. The dependent variable is an indicator for whether the student drops out during that year. The main explanatory variables are the coherence accumulated, semester credit hours accumulated, and GPA in the corresponding year and the preceding years. The main explanatory variables are scaled to their standard deviations in each year. Robust standard errors are reported in parentheses.

Table 6: IV Regressions for Effects of Coherence on Dropout

	Dropout			
	Freshmen	Sophomore	Junior	Senior
	(1)	(2)	(3)	(4)
Coherence - Freshmen	-0.071	-0.023	-0.144	-0.102
	(0.018)	(0.023)	(0.021)	(0.022)
Coherence - Sophomore		0.020	-0.028	-0.023
		(0.026)	(0.021)	(0.019)
Coherence - Junior			-0.087	-0.014
			(0.029)	(0.025)
Coherence - Senior				-0.119
				(0.020)
N	186,445	144,906	136,889	157,275

Notes: This table presents the results of 2SLS regressions for the effects of coherence accumulated on drop out. Coherence accumulation is instrumented with the staggered introduction of DW, as explained in section 3.2. Columns (1) — (4) presents result for freshmen, sophomore, junior, and senior years respectively. Each regression restricts to the sample of students who reach the corresponding year. The dependent variable is an indicator for whether the student drops out during that year. The main explanatory variables is the coherence accumulated, scaled to its standard deviations in each year. Robust standard errors are reported in parentheses.

Table 7: IV Regressions for Effects of Coherence on Dropout — Heterogeneity Results

	Dropout			
	Freshmen	Sophomore	Junior	Senior
	(1)	(2)	(3)	(4)
	A. STEM Majors			
Coherence - Freshmen	-0.095	0.058	-0.157	-0.224
	(0.000)	(0.036)	(0.031)	(0.041)
Coherence - Sophomore		0.100	0.020	-0.081
		(0.033)	(0.027)	(0.033)
Coherence - Junior			-0.207	-0.100
			(0.036)	(0.032)
Coherence - Senior				-0.102
				(0.029)
N	74,775	53,710	48,477	59,948
	B. First Gen Students			
Coherence - Freshmen	-0.046	0.005	-0.124	-0.135
	(0.029)	(0.034)	(0.033)	(0.031)
Coherence - Sophomore		0.030	-0.123	-0.124
		(0.035)	(0.029)	(0.026)
Coherence - Junior			-0.051	-0.036
			(0.042)	(0.033)
Coherence - Senior				-0.137
				(0.027)
N	77,872	53,723	45,798	47,935

Notes: This table presents the results of 2SLS regressions for the effects of coherence accumulated on drop out for the subsamples of students who (a) declared a STEM major in the year of interest; (b) is a first-generation college student. Coherence accumulation is instrumented with the staggered introduction of DW, as explained in section 3.2. Columns (1) — (4) presents result for freshmen, sophomore, junior, and senior years respectively. Each regression restricts to the sample of students who reach the corresponding year. The dependent variable is an indicator for whether the student drops out during that year. The main explanatory variables is the coherence accumulated, scaled to its standard deviations in each year. Robust standard errors are reported in parentheses.

Table 8: Effect of Coherence, Credits, and GPA on Persistence, Associate Degree Attainment, Transfer to 4-Yr College, and Bachelor's Degree Attainment for Community College Students

	Dropout		Associate Degree		Transfer to 4-Yr		Bachelor's Degree	
	1st Year	2nd Year	1st Year	2nd Year	1st Year	2nd Year	1st Year	2nd Year
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coherence - 1st Year	-0.007	0.001	-0.000	-0.008	0.009	0.012	0.009	0.028
	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)	(0.003)	(0.001)	(0.003)
Coherence - 2nd Year		-0.006		-0.009		0.019		0.001
		(0.003)		(0.003)		(0.003)		(0.002)
Semester Credit Hours - 1st Year	-0.057	-0.035	0.041	0.011	0.043	0.035	0.024	0.011
	(0.003)	(0.003)	(0.000)	(0.005)	(0.003)	(0.005)	(0.000)	(0.003)
Semester Credit Hours - 2nd Year		-0.099		0.087		0.084		0.056
		(0.003)		(0.003)		(0.003)		(0.003)
GPA - 1st Year	-0.175	-0.008	0.110	0.015	0.132	0.003	0.074	0.019
	(0.001)	(0.004)	(0.001)	(0.006)	(0.001)	(0.004)	(0.001)	(0.004)
GPA - 2nd Year		-0.182		0.150		0.126		0.072
		(0.003)		(0.003)		(0.003)		(0.003)
N	108,833	27,657	108,833	27,657	108,833	27,657	108,833	27,657

Notes: This table presents the results of regressions in the form of equation (3). Columns (1) — (2) presents results for dropping out without transfer or Associate Degree. Columns (3) — (4) presents results for Associate Degree attainment. Columns (5) — (6) presents results for transferring to a four-year college. Columns (7) — (8) presents results for Bachelor's Degree attainment. Each regression restricts to the sample of students in community colleges who reach the corresponding year. The main explanatory variables are the coherence accumulated, semester credit hours accumualted, and GPA in the corresponding year. The main explanatory variables are scaled to their standard deviations in each year. Robust standard errors are reported in parentheses.

Table 9: Production Functions of Coherence

	Coherence			
	Freshmen	Sophomore	Junior	Senior
	(1)	(2)	(3)	(4)
HS Test Score Decile	0.0040	0.0030	0.0030	0.0014
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Parent Income Decile	0.0014	0.0011	0.0009	0.0004
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
First Gen	-0.0030	-0.0017	-0.0007	-0.0013
	(0.0001)	(0.0001)	(0.0001)	(0.0004)
N	181,146	140,131	133,962	155,491

Notes: This table presents the estimates for β from regressions in the form of equation (7). HS test score is the students' performance in the Texas standardized exams in high school. First Gen is an indicator for first generation college students.

Table 10: Ranking and Production of Coherence and Credits

	Coherence			
	Freshmen	Sophomore	Junior	Senior
+100 in Ranking	-0.0040	-0.0017	-0.0009	0.0015
	(0.0004)	(0.0005)	(0.0004)	(0.0002)
	Semester Credit Hours			
	Freshmen	Sophomore	Junior	Senior
+100 in Ranking	0.080	0.120	0.060	0.070
	(0.007)	(0.008)	(0.001)	(0.009)
N	435,440	381,605	372,127	465,656

Notes: This table presents the estimates for λ from regressions in the form of equation (8). The top panel presents results for the production of coherence while the bottom panel presents results for the production of credit accumulation. Robust standard errors are reported in parentheses.

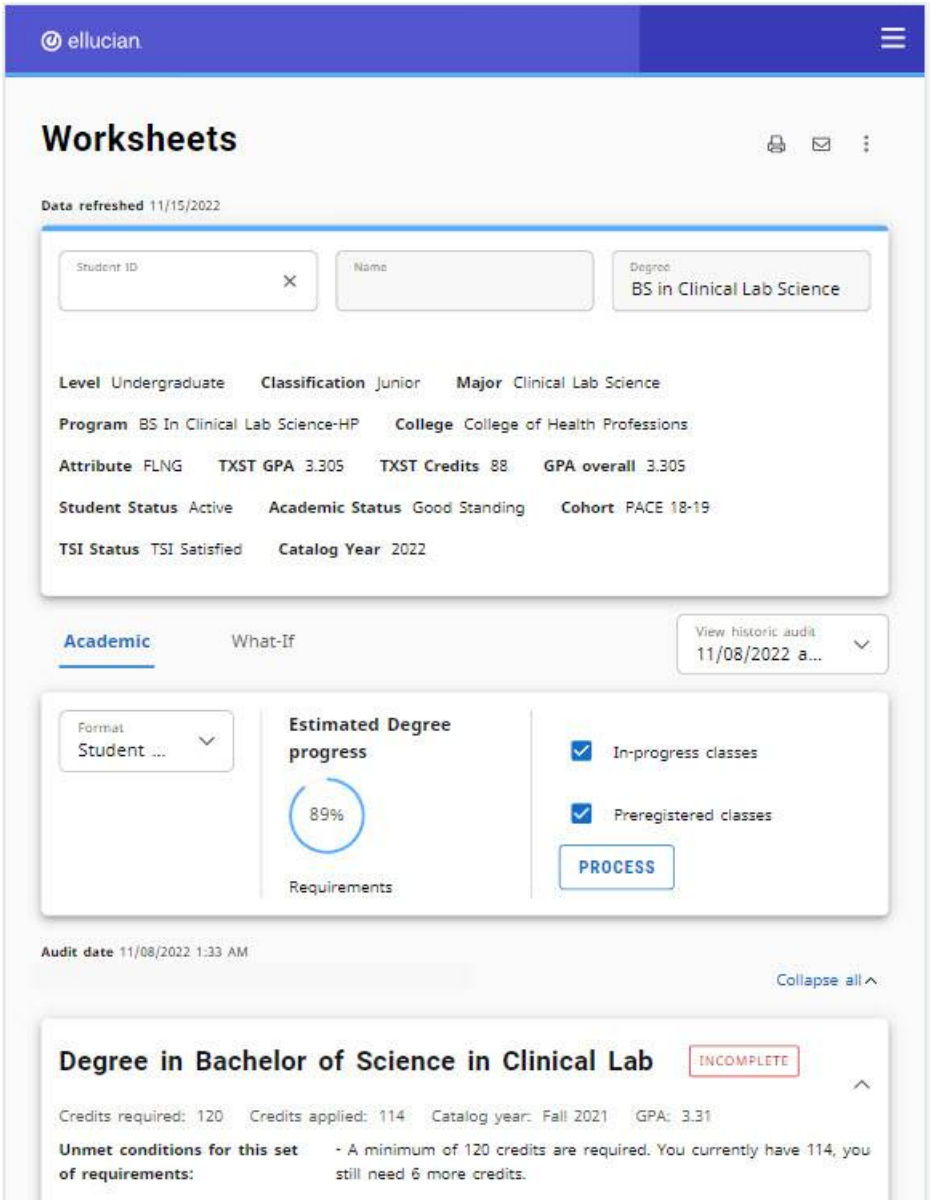


Figure 1: Snapshot of DegreeWorks Interface for Texas State University

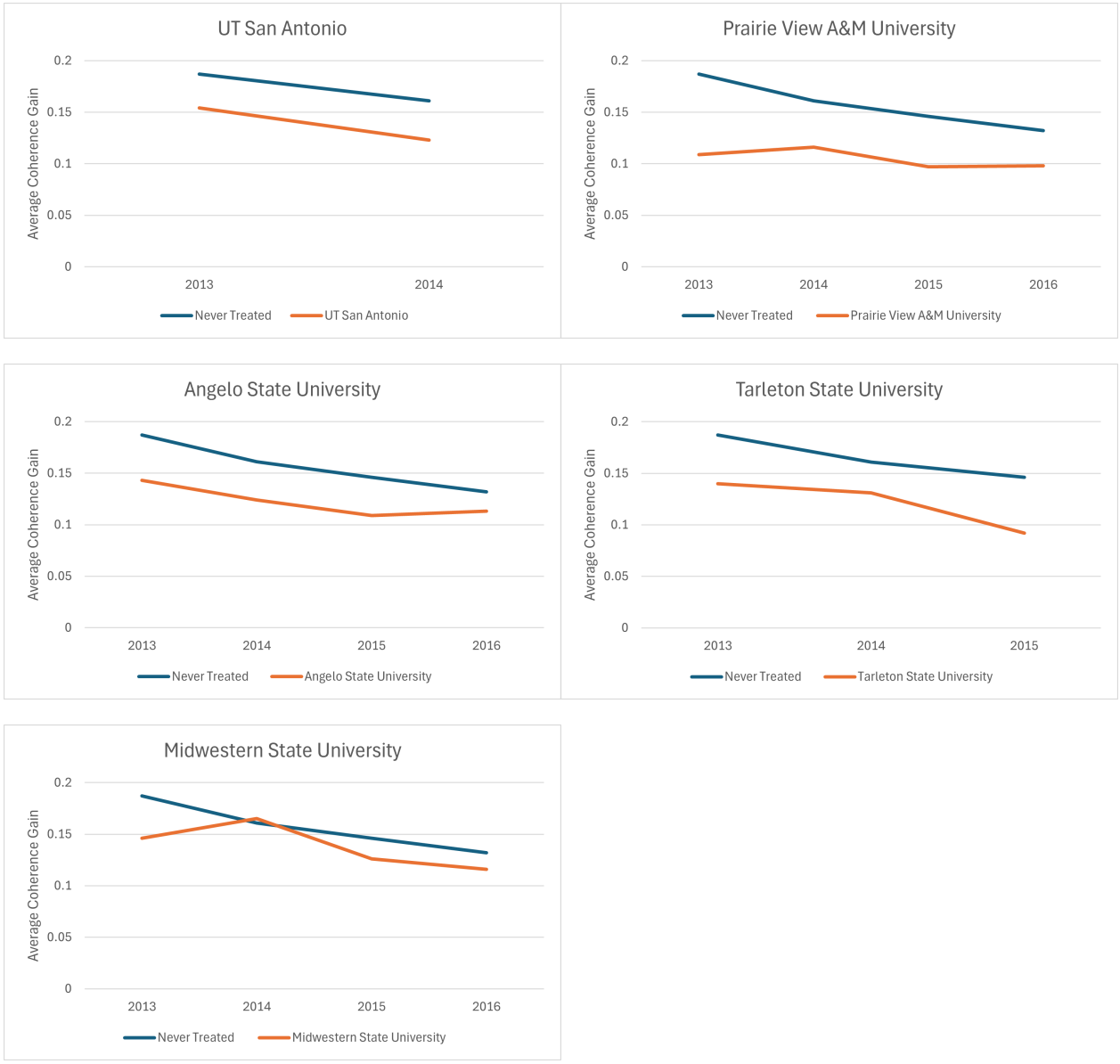


Figure 2: Parallel Trends Checks

Note: the figures present the pre-trends in average coherence gain for students in each treated institution. The last year in each figure is the last pre-treatment year for the corresponding institution.

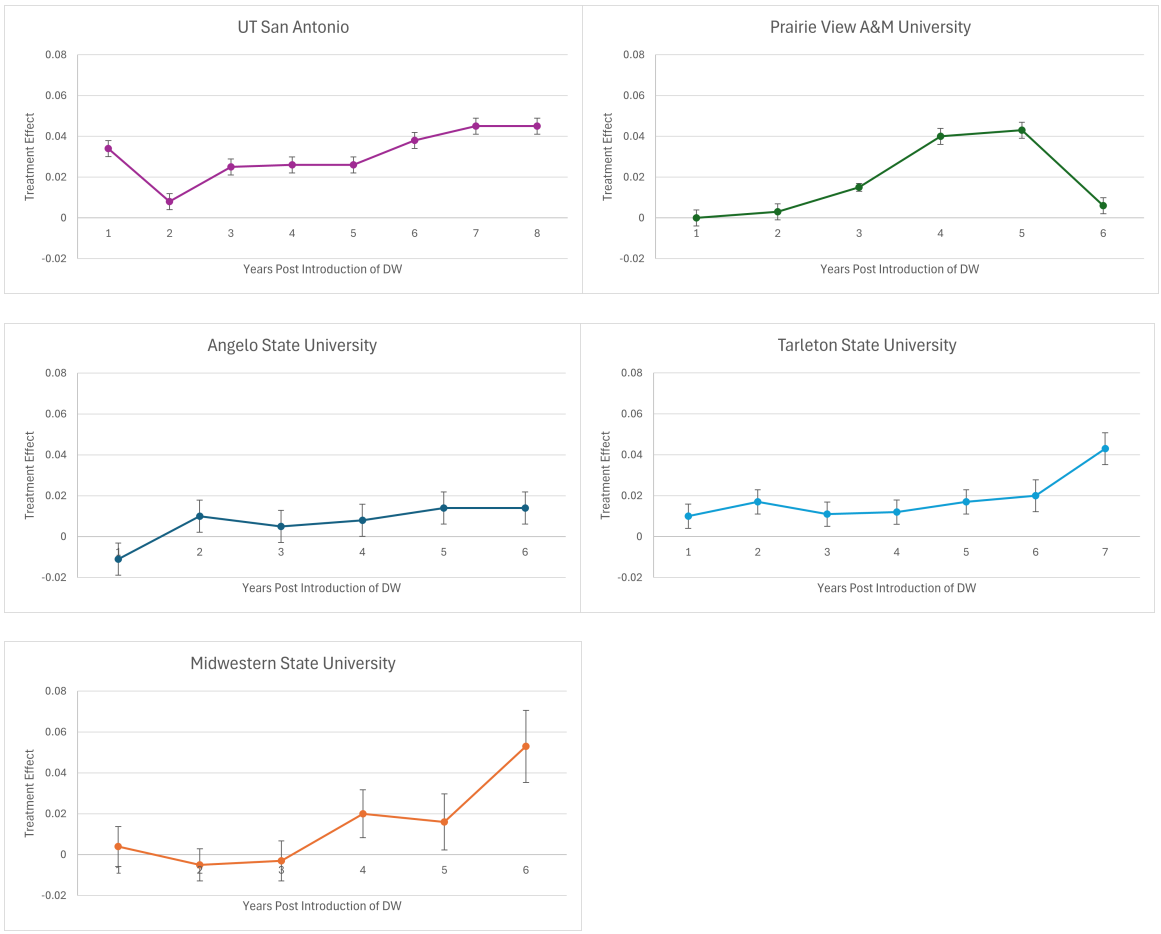


Figure 3: Year × Institution - Specific Treatment Effects of Degree Works on Coherence Accumulation

Note: the figures above present the year×institution-specific treatment effects of Degree Works on coherence accumulation. The horizontal axis is the number of years post-treatment. The number of years post-treatment for which I have treatment effect estimates differ across institutions because Degree Works introduction was staggered. The whiskers are the 95% confidence intervals, calculated with robust standard errors.

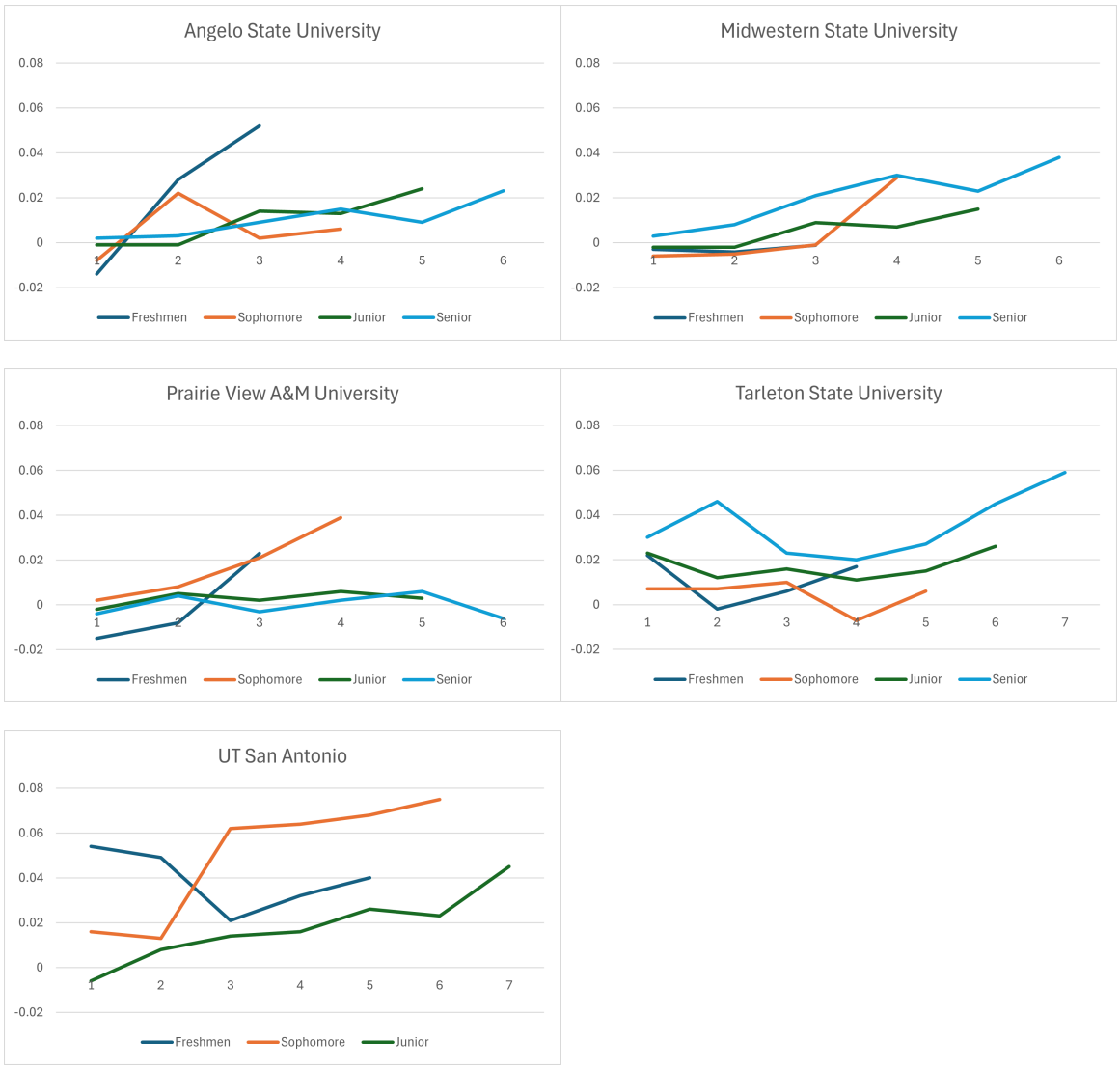


Figure 4: Year × Institution × Academic Standing - Specific Treatment Effects of Degree Works on Coherence

Note: the figures above present the year×institution×academic-standing-specific treatment effects of Degree Works on coherence accumulation. The horizontal axis is the number of years post-treatment. The number of years post-treatment for which I have treatment effect estimates differ across institutions because Degree Works introduction was staggered.

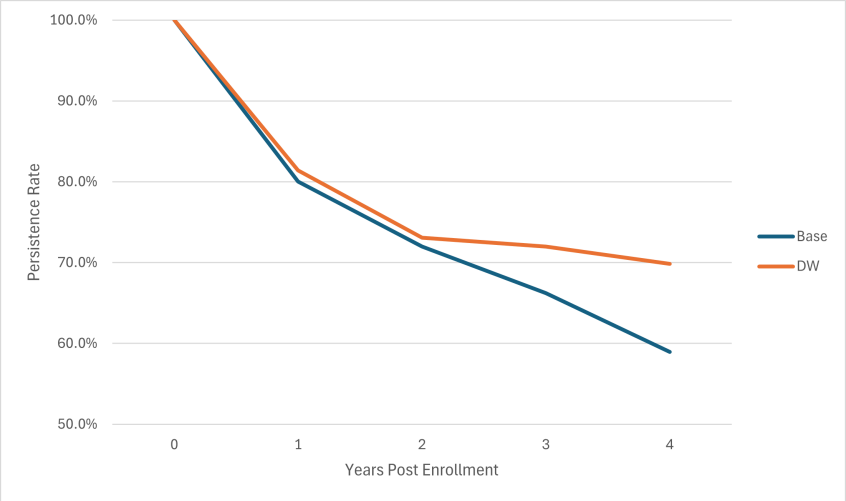


Figure 5: Impact of Coherence Improvement from Degree Works on Persistence

Notes: The figure presents student persistence in the baseline and in the counterfactual world where coherence of students are improved through Degree Works. The impact of Degree Works is calculated as follows. For each of the five treated schools, I take the treatment effect of Degree Works on the coherence of freshmen, sophomores, juniors, and seniors for the last observed treated cohort. I then compute the average treatment effects across schools by weighting with each school's enrollment. Next, I compute the effects on dropout rates for freshmen, sophomores, juniors, and seniors by multiplying the treatment effects on coherence and the estimated effects on dropout rates from the IV regressions. Finally, the counterfactual persistence is calculated with these couterfactual dropout rates. The horizontal axis represents the years after enrollment. The vertical axis represents the share of students that persist after the corresponding year.