A Coherence Measure for the Quality of Students' Course Selections and (Late) College Dropout

Zhengren Zhu*

March 10, 2025

Abstract

I develop a novel measure, *coherence*, for the quality of students' course selections and show that it is closely tied to students' persistence in college. I find that poor course selections is not only associated with higher dropout rate in the current year but also has strong lagging effects on late dropout — a one standard deviation drop in freshmen coherence is associated with 3.6 percentage points higher freshmen dropout rate as well as 2.9, 2.7, and 2.3 percentage points higher sophomore, junior, and senior dropout rate. 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 coherence. Leveraging the staggered introduction of a web-based advising tool in Texas, I find that web-based advising tools can be highly effective in improving the quality of students' course selections. The introduction of the web-based advising tool also provides quasi-experimental variations in coherence, which I use to validate the effects of coherence on dropout rates.

1 Introduction

This paper introduces an easily calculable measure for the quality of students' course selections. 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 institution × major × cohort cell by examining the transcripts of students in each cell that have

*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.

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 the student's major. Following this definition, a student gains coherence when the student makes course selections that are coherent, in the sense that they are logical and contribute to degree completion.[1]

Equipped with the coherence measure, I demonstrate its importance in explaining student persistence. I pay particular attention to *late* college dropout, which is a prevalent, costly, yet understudied type of dropout. While the overall college graduation rate in the US is indeed low (64% college students graduate within 6 years of matriculation), what is perhaps more disturbing is that many students drop out of college when they are close to the finish line. Using administrative data 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. Moreover, these late college dropouts are not unique to unprepared students in non selective institutions — a recent study reports that even highly qualified college students in a highly selective institution drop out after more than three years in college at a non-trivial frequency (Cohodes, Ho, and Robles, 2022). Applying the attrition rates found in the Texas data 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. 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), implying that delaying dropout by one year incurs an additional cost of $72,000 to the society as a whole.[2] As a result, simply shifting the junior-year and senior-year dropouts to dropping out during the sophomore year could save more than $10 billion for the society each year.

OLS estimates suggest that coherence has a strong contemporaneous *and* lagging effect on students' dropout rates. A one standard deviation drop in freshmen coherence is associated with 3.6 percentage points higher freshmen dropout rate as well as 2.9, 2.7, and 2.3 percentage points (pp) higher sophomore, junior, and senior dropout rates. In comparison, credit accumulation and GPA have strong contemporaneous but no lagging effects on dropout rates. One explanation for the lagging effect of coherence is the lack of academic advising. Students observe their accumulation of semester credit hours and GPA directly on their end-of-

[1] Under this definition, coherence will naturally increase as students take more courses and will therefore be confounded with number of credits accumulated. In all analyses, I control for number of credits accumulated to separate quality from quantity. An alternative definition would be to define coherence as the share of courses completed that are degree-required. While this definition would circumvent confounding quality with quantity, it would also penalize students who take additional elective courses that are not part of degree plan. For example, consider two students. In her first semester, student A took Introduction to Microeconomics, Introduction to Macroeconomics, Calculus, and Probability, all of which are required for an economics degree. In comparison, student B also took the four courses student A took, but also took linear algebra, which is not required for an economics degree. The second definition of coherence would penalize student B, which is not sensible.

[2] This calculation assumes that the cost of attendance reflects the direct cost for students and the public combined, and assumes that students who drop out have the same earnings potential regardless of the number of years enrolled

semester transcript reports, and can decide whether they would like to stay enrolled or dropout accordingly. However, the quality of course selections is typically not reported on 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, low quality course selections can fly under the radar, and students who are seemingly close to graduating may actually be missing multiple degree-required courses. This could prompt late dropouts when detected.

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).

To study how policies can improve the quality of students' course selection, I evaluate the impact of recent, staggered, introduction of a web-based advising tool named DegreeWorks (DW) in Texas public universities. Using an event study design that allow for dynamic treatment effects (de Chaisemartin and D'Haultfoeuille, 2023), I find that the web-based advising tool significantly improved students' coherence by 1.7 pp., or 21% of a standard deviation. This results is encouraging because it demonstrates that the quality of course selections can be improved substantially through relatively cheap interventions that do not require significant human capital investment.

The staggered introductions of DW also provide a plausibly exogenous variation in students' course selection coherence. Using the event study design as the first stage, I use instrumental variable (IV) regressions to provide additional validation for the contemporaneous and lagging effects of coherence on students' dropout rates. As will be discussed in Section 4.2, the IV regressions are subject to dynamic sample selection issue. However, this issue is likely minor, and the validation exercise provides additional evidence that the contemporaneous and lagging effects of coherence on dropout rates are likely causal.

This paper contributes to the following four 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 several important factors, 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; Denning et al, 2022). 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. 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 "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 DW provides direct evidence that web-based advising tools can be highly effective in supporting academic advising (Arteaga et al, 2022).

Third, this paper is related to two strands of works that study students' fields of study. First, this paper is related to the 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 the student's 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.

Fourth, this paper makes a data contribution to the broad literature using structural models to study

students' school choices (). Structural work on school choices require the specification of students' non-pecuniary preferences for fields and schooling. Without a convenient measure for students' preferences, prior structural works often make strong assumptions on students idiosyncratic preferences. The coherence measure can be calculated easily and can be incorporated into models as a state variable alongside GPA and credit accumulation. Conditioning on students course selection significantly relaxes the assumptions needed on idiosyncratic preferences.

The remainder of the paper proceeds as follows: Section 2 introduces the Texas administrative data, defines the coherence measure, and explains how the coherence measure is calculated using the Texas administrative data. Section 3 presents empirical results on the relationship between coherence and dropout rates and results on factors affecting coherence. Section 4 evaluates the impact of the introduction of DW on students coherence, and use IV regressions to validate the relationship between coherence and dropout rates. Finally, Section 5 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.[3] 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

[3] 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.

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. In this setting, high-quality course selection is defined as one that is logical and contributes to a student's degree completion. Following this definition, the coherence measure is designed to capture a student's progress towards completing degree-required courseworks and is defined as the share of required courses that the student has completed. Formally,

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

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 the student's coherence in the declared major j^{*} if j^{*} is declared. If the student has not yet declared a major, the student's coherence is defined as the highest coherence among 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, $C^R_{J,M,T}$. Third, I identify courses that more than 30% of $S_{J,M,T}$ took, and label these courses as the set of elective courses, $C^E_{J,M,T}$. 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, $n^E_{J,M,T}$. The tuple $\{C^R_{J,M,T}, C^E_{J,M,T}, n^E_{J,M,T}\}$ 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 $C^R_{J,M,T}$, and at least $n^E_{J,M,T}$ courses in the set $C^E_{J,M,T}$.

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 transferable between institutions. The TCC required courses are taken directly from THECB official websites.

2.3 Validating the Coherence Measure

I first 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.[4] Appendix Table A.1 shows the accuracy rates of the 16 institution $\times$ major pairs I tested. For each pair I audited, more than or close to 90% of empirically identified degree requirements were found on the official degree plan, except for 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.

Next, I demonstrate the importance of the coherence measure in predicting student outcomes. I compare 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 2.1. 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

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

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 to when only credits accumulated and GPA were used.

Repeating the same exercise on community college students (reported in Table 2.2), 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 considerably weaker than in four-year colleges, because many students choose to transfer to a four-year college before attaining an associate's degree.

3 Empirical Specifications and Results

3.1 Empirical Specifications

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:

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

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.[5]

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

[5] This is because the coherence measure is institution-specific and incorporating courses transferred from other institutions complicates the problem significantly due to uncertainty in transferrability of credits. I leave this aspect of calculating coherence to future works.

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.

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

$$\text{Coherence}_{it} = \alpha_0 + \alpha_x X_{it} + \alpha_s S_j + \alpha_m M_{it} + \epsilon_{it} \quad (5)$$

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 (7), 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:

$$\text{Coherence}_{it} = \alpha_0 + \alpha_x X_{it} + \alpha_r \text{rank}_j + \alpha_m M_{it} + \Omega_i + \epsilon_{it} \quad (6)$$

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.[6]

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

3.2 Empirical Results — Effects of Coherence on Dropout Rates

I present the results of regressions following specification of equation (4). The results are presented in columns (1) - (4) of Table 3, and show results separately for students in their freshmen, sophomore, junior, and senior years. Since coherence, semester credit hours, and GPA have different scales, the coefficients are translated to represent the impact of a one standard deviation increase in each of the three factors. Three patterns emerge. First, looking at contemporaneous effects (in the diagonal cells), one can see that increases in all three factors are correlated with a decrease in the likelihood of dropping out in the current semester. Second, GPA appears to have the strongest contemporaneous correlation with dropout (ranging from 7 pp to 22 pp), followed by semester credit hours (ranging from 0.7 pp to 4.4 pp), and then coherence (ranging from 0.9 pp to 1.8 pp). Third, among the three factors, only coherence appears to have a strong lagging effect on later years' dropout rates. 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. For example, holding junior and senior year coherence gain fixed, a one standard deviation increase in freshmen and sophomore coherence gain both decrease drop out in senior year by 1.7 pp. In contrast, the credit accumulation and GPA of freshmen and sophomore year have no meaningful lagging effects on later drop out decisions.

In addition to the average correlation of coherence on dropouts, two heterogeneity results are reported in Table 4. First, I find that the lagging effects of coherence are generally more significant for first generation college students. For example, a one standard deviation improvement in freshmen coherence decreases junior and senior year drop out rates by 1.0 pp and 1.7 pp for the average student but by 1.2 pp and 2.0 pp for first-gen 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. Second, I find that the lagging effects of coherence on late dropouts are more significant for STEM majors. For example, a one standard deviation improvement in freshmen and sophomore coherence decreases senior year drop out rates by 1.7 pp for the average student but by 2.2 pp and 2.1 pp 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.

While interesting, it should be noted that these results are based on OLS and therefore subject to endogeneity concerns. However, based on results from instrumental variables regressions that I present in Section 4, the estimated effects are at least partially causal.

I repeat the regressions from equation (4) on community college students, and the results are presented in Table 5. 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 is negatively correlated with the likelihood of dropping out, although the correlations are weaker than those of semester credit hours and GPA. Columns (3) - (8) break out the correlation with dropout into correlations with associate degree attainment, transfer to four-year colleges, and bachelor's degree attainment. The results suggest that, while coherence accumulation is not correlated with the likelihood of associate degree attainment, it has a large positive correlation with 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 transferable 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 5. 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 for the likelihood of transferring to a four-year college and attaining a bachelor's degree.

3.3 Empirical Results — Factors Effecting Coherence

I now 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 6. Students with higher high-school test scores accumulate coherence faster — a one decile 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. These results suggest that parental resources, both monetary and non-monetary, may be substitutes 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 7. The results show that higher ranked institutions are better at helping students accumulate coherence faster. A drop 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 0.2 pp for freshmen, sophomores, juniors, and seniors, respectively. Comparing to results in Table 6, this is similar to the effects of a one decile increase in high school test score or a 3 decile increase in parental income. In comparison, enrolling in lower ranking institutions actually increases the number of semester credit hours accumulated (bottom panel of Table 7).

Notice that, since the standard deviation of coherence accumulation is around 10 pp, the impacts of the factors above are not large.

4 Effect of DW on Quality of Students' Course Selection

Can policies help improve the quality of students' course selections? Academic advising is perhaps the most common way institutions help students make course selections that are coherent with their degree plan. However, improving academic advising is potentially costly, since it may involve hiring additional advisors or demanding faculty to spend more time on advising. As explained in the next sub-section, DW provides a promising solution to improve academic advising with relatively low cost. I evaluate whether the implementation of DW in Texas public universities actually improved the quality of students' course selection. In addition, I take the introduction of DW as an exogenous shock to students' coherence and use instrumental variables method to validate some of the results from Section 3.

4.1 Institutional Backgrounds on DW

DegreeWorks (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 course works. 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 A.1 shows a snapshot of the DW tool for Texas State University. The example on the figure 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 1 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 implementations of DW as sources of exogenous variations for 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.[7] Combined, these five institutions enroll around 10% of all public university students in Texas.

4.2 Event Study Design and Instrumental Variables Method

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. I follow de Chaisemartin and D'Haultfoeuille (2023), and estimate the treatment effects of DW on students' coherence accumulation using a heterogeneity-robust event-study difference in differences design. The standard two-way fixed effects (TWFE) regression is:

$$\text{coherence}_{it} = \gamma_0 + \gamma_1 X_{it} + \gamma_2 M_{it} + \lambda_j + \delta_t + \gamma d_{jt} + \nu_{it} \tag{7}$$

where λ_j is the institution FE and δ_t is the year FE. d_{jt} is a dummy variable for having DW available and γ is the treatment effect estimated.

The dynamic TWFE regression is as follows:

$$\text{coherence}_{it} = \gamma_0 + \gamma_1 X_{it} + \gamma_2 M_{it} + \lambda_j + \delta_t + \Gamma \mathrm{D}_{jt} + \nu_{it} \tag{8}$$

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

[7] Texas A&M - Texarkana, Texas A&M - San Antonio, Texas Tech, and UT - Rio Grande Valley implemented DW in the 2011-2012 academic year, but since I do not observe any pre-treatment periods for these schools, they also would not contribute to identification.

effects of DW for students in the corresponding $\{j, t\}$ cell. I also estimate placebo estimators, that compare outcomes of the third and second year before treatment to the last year pre-treatment. These estimates will be used to test the parallel trends and no-anticipation assumptions.

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, and the OLS estimates presented in Section 3.2 might therefore be biased. 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 as a source of exogenous shift in students' coherence. Since the dynamic TWFE estimated using equation (8) is not significantly different from the standard TWFE estimates from equation (7), I use equation (7) as the first stage for IV estimations of equations (3), with d_{jt} used as excluded instrument.

The exclusion restriction for the IV method is that DW only affects dropout through students' course selection. There are two reasons why the exclusion restriction might not be satisfied. First, DW could affect other intermediate outcomes, such as credits accumulated. While DW is designed to only provide automatic degree audit, I cannot prove that institutions did not simultaneously upgrade the course registration system when implementing DW. I mitigate this concern by controlling for students' semester GPA and credit hours accumulated. Second, DW could have an information effect. Specifically, conditional on the same course selection coherence, DW makes it easier for students to observe their progress towards degree, which may impact dropout decisions. I cannot rule out this potential mechanism. However, this effect would likely bias the estimate negatively — since students tend to be overconfident about their ability to complete college under imperfect information (Stinebrickner and Stinebrickner, 2014), the information effect from DW would increase, not decrease, dropout rates.

Finally, it is important to note that the IV method is meant to help validate the causal link between coherence and dropout and not meant to supersede the OLS estimates reported in Section 3.2. Apart from the potential information effect mentioned in the previous paragraph, the IV method cannot be used to credibly estimate the model with lagged coherence levels in equation (4). First, equation (4) has two to four endogenous variables when looking at sophomore to senior dropout rates. The staggered introduction of DW only provides one instrument. While interacting the staggered introduction with students' cohorts can provide sufficient number of instruments, this may lead to concerns of weak instruments. Second, there is a dynamic sample selection issue. For example, the effect of coherence on freshmen dropout rate will lead to endogenous sample selection when examining the dropout rate of sophomores. There is no clear way to address this issue.

4.3 Event Study Results

The estimates of the dynamic treatment effects of DW on coherence are reported in and . Both show significant improvement in coherence after the introduction of DW. The average increase in coherence gain per semester is 1.7 percentage points, which is 21% of one standard deviation. The effect seems to increase over time, with the effect in the fourth year after introduction reaching 2.4 percentage points. show that there is no significant placebo treatment effects, validating the parallel trends assumption. The p-value for the test of joint nullity of the placebos is 0.30. The first column of reports the estimates from a standard TWFE regression. The result shows a treatment effect of 2.1 percentage points, which is close to the average dynamic treatment effect computed.

also reports the treatment effect estimates for two sub-samples — STEM majors and first generation college students. The average treatment effects, 1.7 percentage points and 1.5 percentage points, are close to those estimated with the full sample.

compares the OLS and IV estimates of equation (3). The OLS estimate show that, on average, a one standard deviation improvement in coherence decreases dropout rate by 1.3 percentage points. The IV estimate show that a one standard deviation increase in coherence decreases dropout rate by 3.0 percentage points, though the standard errors are larger. The F statistics for the first stage instrument is 557.9. Conditional on all caveats discussed in Section 4.2, I use IV method to estimate equation (4) with the interaction of students' cohort and the introduction of DW as exogenous instruments. The results (unreported) show the same qualitative patterns as the results reported in .

5 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 and 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 require 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 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.

Arteaga, Felipe, Adam J. Kapor, Christopher A. Neilson, and Seth D. Zimmerman. 2022. "Smart Matching Platforms and Heterogeneous Belief in Centralized School Choice." Quarterly Journal of Economics 137 (3): 1791-1848.

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.

Denning, Jeffrey T., Eric R. Eide, Kevin J. Mumford, Richard W. Patterson, and Merrill Warnick. 2022. "Why Have College Completion Rates Increased?" *American Economic Journal: Applied Economics* 14 (3): 1-29.

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.

Stinbrickner, Ralph and Todd Stinebrickner. 2014. "Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model." *Journal of Labor Economics* 32(3): 601-644.

Tables and Figures

Table 1.1: 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 1.2: 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 2: Effects of Lagging and Current Coherence, Credits, and GPA on Dropout

	Dropout			
	Freshmen	Sophomore	Junior	Senior
	(1)	(2)	(3)	(4)
Coherence - Freshmen	-0.014	-0.007	-0.010	-0.017
	(0.001)	(0.001)	(0.001)	(0.007)
Coherence - Sophomore	-	-0.009	-0.010	-0.017
		(0.001)	(0.001)	(0.001)
Coherence - Junior	-	-	-0.016	-0.016
			(0.001)	(0.001)
Coherence - Senior	-	-	-	-0.018
				(0.001)
Semester Credit Hours - Freshmen	-0.039	0.001	0.003	0.004
	(0.001)	(0.001)	(0.001)	(0.001)
Semester Credit Hours - Sophomore	-	-0.044	0.001	-0.003
		(0.001)	(0.001)	(0.001)
Semester Credit Hours - Junior	-	-	-0.044	-0.011
			(0.001)	(0.001)
Semester Credit Hours - Senior	-	-	-	-0.007
				(0.001)
GPA - Freshmen	-0.219	0.019	0.007	0.003
	(0.001)	(0.001)	(0.001)	(0.001)
GPA - Sophomore	-	-0.096	0.004	-0.004
		(0.001)	(0.001)	(0.001)
GPA - Junior	-	-	-0.070	-0.012
			(0.001)	(0.001)
GPA - Senior	-	-	-	-0.067
				(0.001)
N	170,771	126,514	112,711	130,503

Notes: This table presents the results of OLS regressions in the form of equation (4). Columns (1) — (4) present results 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. So the coefficients should be interpreted as the effect of a one standard deviation increase in each variable on dropout rate. Robust standard errors are reported in parentheses.

Table 3: Effects of Lagging and Current Coherence on Dropout — Heterogeneity Results

	Dropout			
	Freshmen	Sophomore	Junior	Senior
	(1)	(2)	(3)	(4)
	A. First Generation College Students			
Coherence - Freshmen	-0.016	-0.008	-0.012	-0.020
	(0.002)	(0.001)	(0.002)	(0.001)
Coherence - Sophomore	-	-0.012	-0.013	-0.022
		(0.002)	(0.001)	(0.001)
Coherence - Junior	-	-	-0.019	-0.021
			(0.002)	(0.001)
Coherence - Senior	-	-	-	-0.018
				(0.002)
N	74,530	51,662	43,853	47,816
	B. STEM Majors			
Coherence - Freshmen	-0.021	-0.014	-0.010	-0.022
	(0.002)	(0.001)	(0.001)	(0.001)
Coherence - Sophomore	-	-0.017	-0.010	-0.021
		(0.001)	(0.000)	(0.001)
Coherence - Junior	-	-	-0.020	-0.024
			(0.001)	(0.001)
Coherence - Senior	-	-	-	-0.030
				(0.002)
N	66,228	47,523	44,034	56,981

Notes: This table presents the results of OLS regressions that estimate equation (4) for the subsamples of students who (a) is a first-generation college student; or (b) declared a STEM major in the year of interest. Columns (1) — (4) present 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 in the corresponding year and the preceding years, scaled to its standard deviations. The regression also controls for current and lagging semester credit hours and GPA. Robust standard errors are reported in parentheses.

Table 4: 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 accumulated, 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 5: 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 6: 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), controlling for students' application and admission portfolio. 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.

Table 7: Institutions with DegreeWorks and Year of Implementation

Institution	Academic Year Implemented DW
Texas A&M - Commerce	2010-2011
Texas A&M - Corpus Christi	2010-2011
Sam Houston University	2011-2012
Texas A&M - Kingsville	2011-2012
Texas State University	2011-2012
Texas A&M - Texarkana	2011-2012
Texas A&M - San Antonio	2011-2012
Texas Tech University	2011-2012
UT - Rio Grande Valley	2011-2012
UT - San Antonio	**2014-2015**
Tarleton State University	**2015-2016**
Angelo State University	**2016-2017**
Midwestern State University	**2016-2017**
Prairie View A&M University	**2016-2017**

Notes: this table presents the list of Texas public 4-year institutions that have implemented DegreeWorks, as well as the academic year when the tool was implemented. The information on the year implemented were acquired from the institutions' registrar's office if not directly found on their websites. The five institutions that are the main source of exogenous variation in the event study are highlighted in bold.

Table 8: Event Study Estimates — Effects of Introduction of DW on Coherence

	Coherence			
	TWFE	Dynamic TWFE		
	All Students	All Students	First Gen Students	STEM Majors
	(1)	(2)	(3)	(4)
Average Treatment Effect of DW	0.021	0.017	0.017	0.015
	(0.007)	(0.008)	(0.008)	(0.008)
N	562,274	562,274	226,121	221,954

Notes: this table presents the average treatment effects of the introduction of DW on coherence gain per semester. Column 1 uses standard TWFE (equation (7)), and Columns 2 - 4 use dynamic TWFE (equation (8)) following de Chaisemartin and D'Haultfoeuille (2023). Columns 1 and 2 use the full sample of students. Columns 3 and 4 restrict to first generation college students and STEM majors, respectively. All specifications control for basic controls, including gender, ethnicity, high school standardized test scores, and parental income, as well as GPA, semester credit hours, and academic standing. Standard errors, reported in parentheses, are clustered at school level and calculated using wild bootstrap.

Table 9: Comparison between OLS and IV

	Drop Out Rate	
	OLS	IV
	(1)	(2)
Coherence	-0.013	-0.030
	(0.000)	(0.010)
Semester Credit Hours	-0.017	-0.012
	(0.000)	(0.002)
GPA	-0.101	-0.097
	(0.001)	(0.002)
N	562,274	562,274

Notes: this table presents the results of OLS regressions that estimate equation (3) — column 1 — and IV regressions that estimate equation (3) using introduction of DW as instrument for coherence — column 2. The dependent variable is an indicator for whether the student drops out. The main explanatory variables are the coherence accumulated in the corresponding year. The regression also controls for current semester credit hours and GPA. All specifications control for basic controls, including gender, ethnicity, high school standardized test scores, and parental income. Robust standard errors are reported in parentheses.

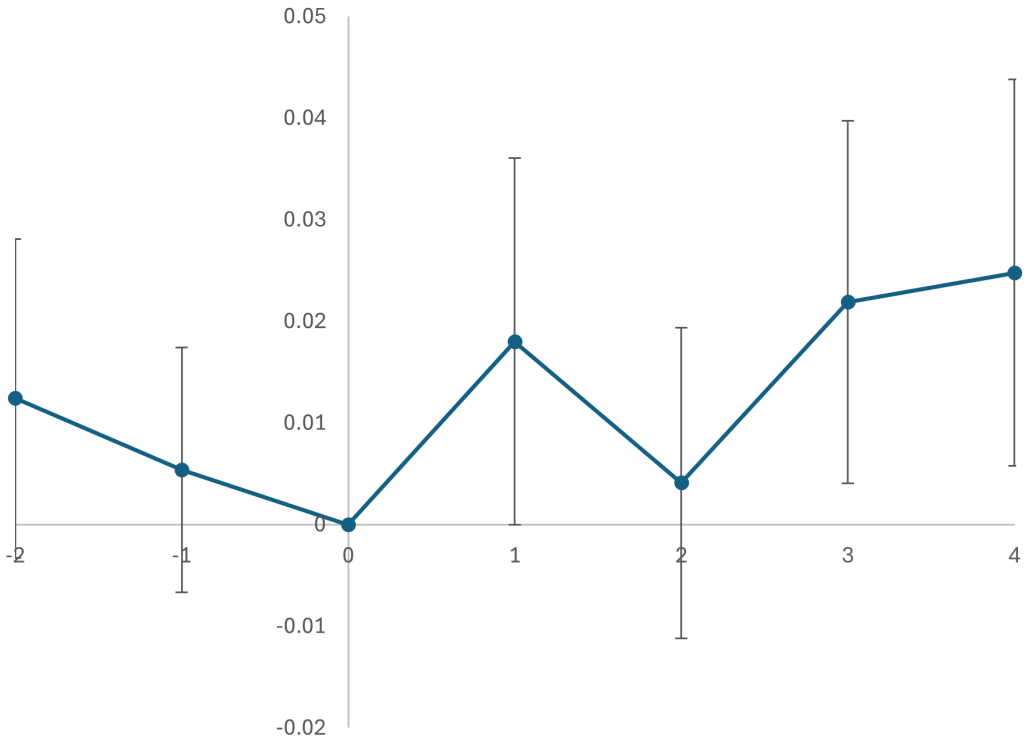


Figure 1: Treatment Effect from Last Period Before Treatment (T=0)

Note: this figure presents the estimated average (placebo) treatment effects of the introduction of DW, relative to the last period before treatment (T=0). The estimates are on the full sample of students and use dynamic TWFE (equation (8)) following de Chaisemartin and D'Haultfoeuille (2023). 95% confidence intervals are shown in wicks. The estimation controls for basic controls, including gender, ethnicity, high school standardized test scores, and parental income, as well as GPA, semester credit hours, and academic standing. Standard errors are clustered at school level and calculated using wild bootstrap.

Appendix: Additional Tables and Figures

Table A.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.

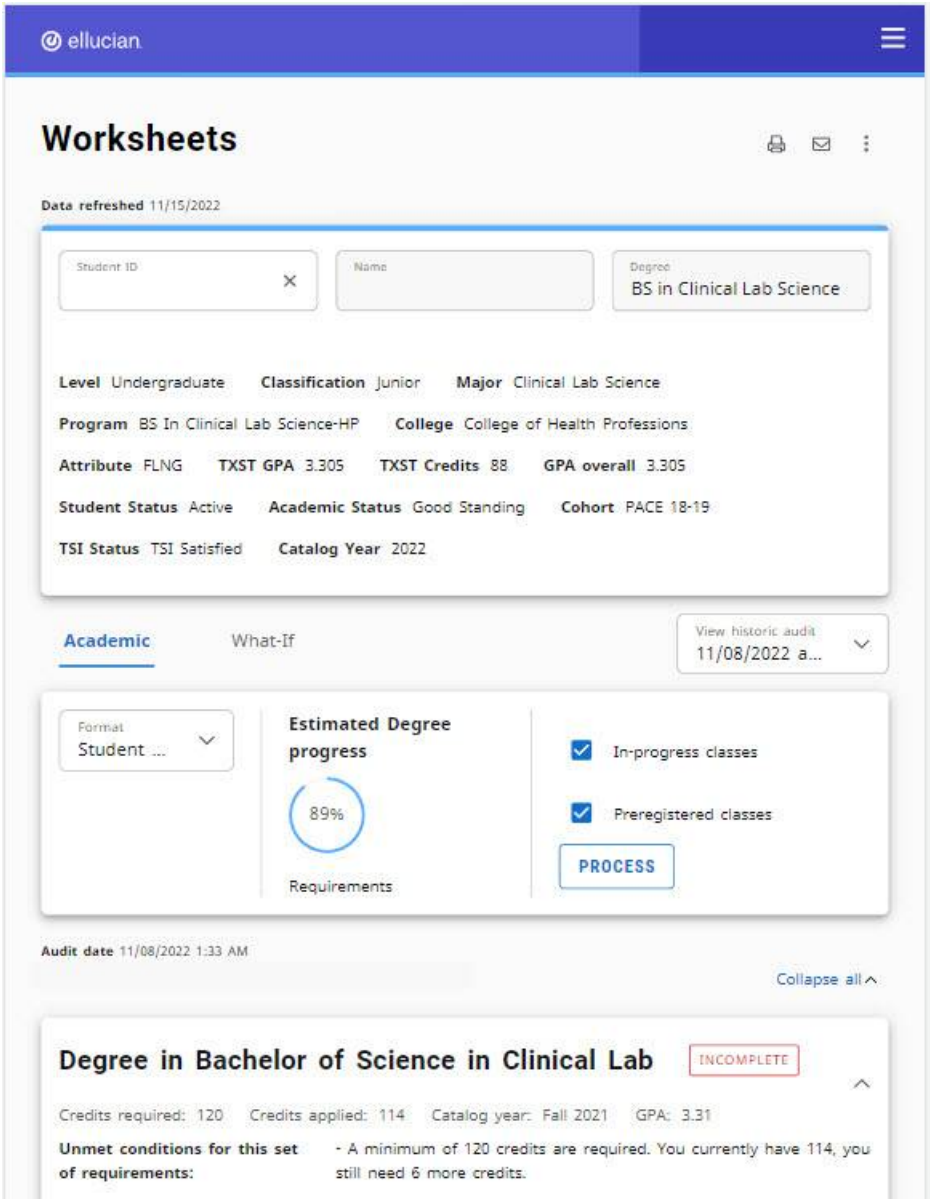


Figure A.1: Snapshot of DegreeWorks Interface for Texas State University

Notes: this figure is a snapshot of Degree Work's functionality at Texas State University. The snapshot was taken directly from Texas State University's website.