Frontier Knowledge in College and Student Success*

Barbara Biasi† Song Ma‡

May 11, 2026

Abstract

We study whether exposure to frontier knowledge in college affects student outcomes. Combining 459,415 syllabi from seven Texas public universities with 107 million publications and linked student records, we measure each course's proximity to recent versus older research in its field. Exploiting syllabus updates unobserved at enrollment, we find that frontier exposure increases completion, GPA, graduate-school attendance, and earnings, and reduces time-to-degree. Completion, GPA, and progression gains are broad, while graduate-school and earnings returns are larger for students with stronger preparation and family resources. The evidence suggests two mechanisms: frontier content keeps students engaged, and sustained exposure builds labor-market skills.

JEL Classification: I23, I24, I26, J24, O33

Keywords: Higher Education, Frontier Knowledge, Course Content, Knowledge Diffusion, Syllabi, Text Analysis, Human Capital, Inequality

*The conclusions of this research do not necessarily reflect the opinion or official position of the Texas Education Research Center, the Texas Education Agency, the Texas Higher Education Coordinating Board, the Texas Workforce Commission, or the State of Texas. We thank Jaime Arellano-Bover, David Autor, Joe Altonji, Raj Chetty, Ben Sprung-Keyser, Zhengren Zhu, Seth Zimmerman, and seminar participants at various institutions and conferences for helpful comments. Meghna Baskar, Xugan Chen, Yajie Luo, and Xinhui Yu provided outstanding research assistance. We thank the Yale Tobin Center for Economic Policy, The Broad Center at the Yale School of Management, the Yale School of Management Dean's Fund, the Yale Center for Research Computing, Yale University Library, and the Yale International Center for Finance for research support. All errors are our own.

†Yale School of Management and NBER, barbara.biasi@yale.edu;

‡Yale School of Management and NBER, song.ma@yale.edu.

1 Introduction

Colleges and universities are central to the production and diffusion of frontier knowledge—the most current ideas, methods, applications, and evidence in a field. Through their faculty, they produce a large share of frontier research (Andrews, 2023); through their courses, they have the potential to diffuse that knowledge to hundreds of thousands of students each year. This dual role matters because frontier knowledge is a well-established engine of economic growth (Romer, 1986): its production generates the ideas that fuel innovation (Jones, 2009; Iaria et al., 2018), and its diffusion to workers and entrepreneurs transforms those ideas into economic value (Goldin and Katz, 2010; Acemoglu and Autor, 2011). In the classroom, frontier knowledge can appear in the topics and problems a course emphasizes, the contemporary settings it uses to motivate core concepts, and the evidence and methods it introduces.

Yet whether college actually disseminates frontier knowledge to students, and whether such dissemination matters for student outcomes, is ex ante unclear. First, the design of course content is largely delegated to instructors, and even within the same institution and major, syllabi can differ sharply in how much frontier knowledge they incorporate.[1] Second, the effects of frontier knowledge exposure on students are not obvious. Bringing courses closer to the frontier may increase engagement and learning by making material more relevant and by building skills valued in the labor market, but it may also crowd out foundational knowledge or raise difficulty, potentially harming students on fragile completion margins. Answering these questions is challenging. It requires course-level data on what is actually taught, which typically are unavailable in standard administrative records. Identifying causal effects additionally requires linking these data to students' course-taking histories and outcomes and addressing students' selection of courses.

This paper addresses these challenges with newly collected data and a novel empirical strategy. Applying text analysis to 459,415 syllabi from seven major public universities in Texas and to a corpus of over 107 million academic publications, we construct each course's *frontier knowledge proximity*, the degree to which a course's content reflects the current frontier of its field. Courses have higher proximity when they are organized around the questions, applications, evidence, and methods that characterize that frontier.[2] We link these syllabi to individual administrative records covering over 200,000 undergraduate and 40,000 graduate students reporting positive unemploy-

[1] An instructor's right to determine course content, teaching methods, and assessment strategies is grounded in principles of academic freedom, as articulated by the American Association of University Professors (AAUP) in its 1940 Statement of Principles on Academic Freedom and Tenure (American Association of University Professors, 1940).

[2] Studies that have used recent academic publications to capture the research frontier include Iaria et al. (2018) and Angrist et al. (2017).

ment insurance-covered earnings within Texas and exploit an institutional feature of Texas course registration for causal identification: students enrolling in a course can observe the *previous* term's syllabus, but not the updated version. Changes in a course's frontier proximity from one term to the next are therefore unobserved by students at the time of enrollment. Conditional on lagged proximity and instructor identity, these changes provide plausibly quasi-random variation in frontier exposure, which we use to estimate causal effects on educational attainment, academic performance, graduate-school enrollment, and labor-market earnings.

We measure frontier knowledge proximity as the ratio of a syllabus's average textual similarity to recent research publications to its similarity to older ones, computed using a backward-looking term-frequency weighting scheme (Kelly et al., 2021) that suppresses ubiquitous foundational vocabulary and amplifies language distinctive of the current frontier. A higher score therefore identifies courses whose language tracks the contemporary edge of their field rather than its established canon; because the measure is a ratio, differences in syllabus length and verbosity largely cancel. Four validation checks confirm that the measure captures what it is designed to capture: an LLM-based content assessment ranks matched syllabi in the same direction as our measure in 74% of pairs; a simulation that progressively substitutes newer for older content terms raises proximity monotonically; syllabi citing more recent references exhibit higher proximity; and proximity rises monotonically with course level, from lower-undergraduate to graduate.

Applying this metric to our Texas syllabi, we first document the extent of variation in frontier knowledge coverage across courses. The differences are large. Moving a syllabus from the 25th to the 75th percentile of the proximity distribution is equivalent to replacing roughly 30% of its content with newer knowledge. Strikingly, instructors and courses explain the largest shares of this variation (21% and 26%, respectively, from a Shapley-Owen variance decomposition), while differences across schools explain only 0.3% and differences across fields explain 5%. Even within the same institution, field, and course level, students face substantially different exposures to frontier knowledge depending on which courses they enroll in. This pattern motivates the central causal question of the paper: does frontier knowledge exposure affect student outcomes?

To answer that question, we exploit a key feature of the Texas course-registration process: when students register, they can observe the prior term's syllabus and the instructor's name, but not the updated content. Conditional on lagged course proximity and instructor identity, within-course updates in proximity are therefore unobserved by students at enrollment, rendering them conditionally exogenous. We implement this strategy by augmenting our outcome equation with con-

trols for each student's average lagged course proximity and with instructor fixed effects. We also absorb school-by-major-by-cohort fixed effects to account for the menu of courses available in each program and cohort. The resulting identifying variation isolates within-program, within-instructor changes in course proximity that students are unlikely to anticipate at registration, and it corresponds to roughly one third of the overall variation in proximity across all courses. Placebo tests estimating the effects of future course proximity on a student's current rank in the course yield coefficients indistinguishable from zero, supporting the conditional-independence assumption underlying our design.

We first estimate the average effects of frontier exposure on educational outcomes. Among undergraduate students, a one-standard-deviation (SD) increase in average course proximity raises bachelor's degree completion by 4.8 percentage points, reduces time-to-degree by 0.21 years, raises GPA by 0.12 grade points, and increases graduate-school attendance within Texas by 1.4 percentage points. Frontier exposure also benefits graduate students, although these effects are concentrated among Master's students. The academic gains appear to reflect learning that carries across courses: exposure to frontier knowledge in earlier coursework predicts stronger performance in later coursework, with a one-SD increase in prior proximity raising a student's rank in subsequent courses by 0.4 percentile points.

We then turn to labor-market outcomes. A one-SD increase in frontier exposure raises earnings by 2.1% for undergraduates over the first six years after predicted graduation and by 4.5% for graduate students, with gains largely concentrated in the first three post-graduation years. At the undergraduate level, frontier exposure raises earnings mainly by reducing downside risk: it lowers the probability of bottom-quartile earnings by 2.1 percentage points, while leaving top-quartile earnings essentially unchanged. Among graduate students, exposure reduces bottom-quartile earnings risk and increases the likelihood of top-quartile earnings. Most of the earnings gains appear to arise within industries rather than through sorting into higher-paying sectors.

The pattern of effects points to two complementary mechanisms. Frontier exposure may strengthen engagement, motivation, and academic momentum, especially when students are still on fragile persistence margins; it may also build skills whose returns accumulate across courses and appear in later academic and labor-market outcomes.

We probe these margins in two ways. First, we estimate effects separately by year of exposure. Effects on degree completion and GPA are largest when exposure occurs in the first year of college, consistent with an engagement channel operating early in students' trajectories. Effects on

time-to-degree, graduate-school attendance, and earnings are more sustained throughout the curriculum. Second, we compare broad exposure across the transcript with exposure concentrated in a small number of high-proximity courses. Degree completion is affected by both median and 90th-percentile proximity, consistent with isolated high-proximity courses strengthening engagement that affects fragile margins. Median proximity, by contrast, predicts faster degree progress, graduate-school attendance, and earnings more strongly, suggesting that higher-threshold outcomes depend on cumulative skill accumulation rather than on isolated frontier-oriented courses alone.

Heterogeneity by baseline ability shows how these mechanisms map into different outcome margins across students. Across ability groups, frontier exposure improves degree completion, GPA, and time-to-degree, consistent with broad gains in engagement, learning, and academic momentum. Students without reported test scores—a group composed largely of transfer students from community colleges—experience large gains in degree completion and academic progression. Yet gains in graduate-school attendance, mean earnings, and top-quartile earnings rise more strongly with baseline ability, suggesting that converting frontier exposure into higher-threshold skill returns is ability-complementary.

Heterogeneity by family income tells a parallel story. Frontier exposure improves degree completion and time-to-degree across the income distribution, consistent with frontier content strengthening academic progression broadly. However, students from the bottom income quartile see no detectable increase in graduate-school attendance. Graduate-school effects instead emerge above the bottom income quartile and are largest for students from the highest-income families; similarly, earnings point estimates are positive across the income distribution but larger in the upper half. Thus, frontier exposure improves engagement and attainment broadly, while the conversion of frontier-acquired skills into graduate-school progression and higher earnings appears more resource-dependent.

The contrast between attainment gains and skill returns is especially useful for interpreting the earnings effects. If frontier exposure raises earnings by strengthening persistence and academic progression, the earnings premium should be largely absorbed by observed educational outcomes. Earnings gains that remain after conditioning on those outcomes, by contrast, point to returns not captured by measured attainment and performance. We use a descriptive mediation exercise to support this interpretation. In the full undergraduate sample, controlling for graduation, GPA, time-to-degree, and graduate-school attendance reduces the estimated earnings effect from 2.1% to

a small and statistically insignificant value. The attenuation is especially pronounced for lower-income students, students without reported test scores, and students in the bottom half of the observed ability distribution, suggesting that their earnings gains operate mainly through improved attainment and progression. For higher-ability students, some of the earnings premium remains after conditioning on these educational outcomes, consistent with additional labor-market returns to frontier-acquired skills. Taken together, the results imply that frontier knowledge strengthens completion and academic progression broadly, while more advanced educational and labor-market returns appear complementary with the resources and preparation students bring to college.

In sum, our results document substantial heterogeneity in the presence of frontier knowledge across university courses and show that these differences carry meaningful consequences for student outcomes. The benefits of frontier exposure are broad across educational outcomes and particularly pronounced on fragile completion margins. At the same time, the ability to translate frontier exposure into graduate-school progression and earnings is stronger among students with greater pre-college resources. This pattern reveals both the equity potential and the equity limits of frontier knowledge diffusion.

Related literature This paper contributes to several strands of the literature. Most closely related is a body of research on heterogeneity in human capital production, focusing primarily on differences across majors (Altonji et al., 2012; Deming and Noray, 2020), institutional selectivity (Hoxby, 1998; Dale and Krueger, 2014; Mountjoy and Hickman, 2021), and the skill content of college programs (Hemelt et al., 2023; Li et al., 2021). We depart from this literature in three ways: we examine proximity to the knowledge frontier—a previously unmeasured dimension of course content—rather than field of study or institutional type; we document heterogeneity within the same institution and major rather than across them; and we develop a research design that exploits the information structure of Texas course registration to isolate exogenous variation in the content individual students encounter.

Our work also relates to the literature on the causal effects of colleges on students, pioneered by Dale and Krueger (2002). More recent contributions (e.g., Cunha and Miller, 2014; Hoxby and Bulman, 2015; Mountjoy and Hickman, 2021) employ a "school value-added" framework to quantify the causal impact of attending a given institution on outcomes such as earnings. While valuable for ranking schools by the returns they generate, these estimates are difficult to translate into policy guidance: they treat institutional quality as a black box, offering no information on which dimensions of instruction drive positive returns, and they are backward-looking by con-

struction—becoming available only years after the relevant instruction has occurred. By linking student outcomes directly to observed course content, our approach can identify which features of a program help students most, information that can inform curriculum design with a shorter delay.

We also contribute to evidence that the content of education, not only its quantity or institutional setting, shapes students' educational and labor-market success. Much of this evidence comes from primary and secondary education, where curriculum reforms have been shown to affect political and economic attitudes (Cantoni et al., 2017), scientific beliefs and occupational choices (Arold, 2024), academic performance (Dee and Penner, 2017), and later earnings (Joensen and Nielsen, 2009). Evidence from higher education is scarcer. Arteaga (2018) shows that a reform reducing required coursework at a selective university lowered wages, demonstrating that the quantity of college coursework can affect labor-market outcomes. We complement this evidence by showing that the content of coursework matters as well. Using syllabi, we isolate a particular dimension of content and show that exposure to it, even within a given institution and major, affects degree completion, graduate-school enrollment, and earnings, at least in part through a skill-building channel. Our mediation results reinforce this interpretation: among more-advantaged students, part of the earnings premium remains after conditioning on graduation, GPA, time-to-degree, and graduate enrollment. That pattern adds to existing evidence that measured schooling returns largely reflect human capital rather than signaling (Aryal et al., 2022).

Finally, our results speak to the broader literature on how access to existing frontier knowledge fosters the creation of new ideas and innovation (Moser and Voena, 2012; Williams, 2013; Galasso and Schankerman, 2015; Iaria et al., 2018; Biasi and Moser, 2021).[3] Educational institutions are frequently cited for their role in disseminating frontier knowledge, particularly in STEM fields (Baumol, 2005; Toivanen and Väänänen, 2016; Bianchi and Giorcelli, 2020; Akcigit et al., 2025). We extend this work by locating the source of frontier exposure within observed course content rather than attributing it to institutions or fields as a whole, and by tracing how that exposure shapes educational attainment and labor-market outcomes across the socioeconomic and ability distribution.

2 Data

Our empirical analysis combines several distinct data sources: course syllabi, academic publications, individual-level student demographics, academic records, and earnings. We provide addi-

[3] Moser and Voena (2012), Williams (2013), and Galasso and Schankerman (2015) show how, in various settings, easier access to pre-existing patents fosters the creation of new patents. Similarly, Iaria et al. (2018) show that reduced scientific cooperation due to World War II leads to a slow-down in the production of new science, and Biasi and Moser (2021) show that a decline in the cost of accessing frontier knowledge in books leads to an increase in the diffusion of those books.

tional details on the construction of the final dataset in Appendix B and Appendix C.

2.1 Course Syllabi

Our syllabi database covers the majority of courses taught at seven major public universities in Texas: Stephen F. Austin State University (starting in 2009), Sam Houston State University (starting in 2011), Texas A&M University (starting in 2013), University of Houston-Clear Lake (starting in 2010), University of Texas at Austin (starting in 2011), University of Texas at Dallas (starting in 2005), and West Texas A&M University (starting in 2013). We collected these syllabi directly from each university's website, for a total of 459,415 documents corresponding to 27,872 courses taught through 2022.[4] These syllabi represent 51% of all courses offered at these institutions during our analysis period (see Appendix Table C1 for a comparison of matched and unmatched courses). We converted the downloaded PDF documents into machine-readable text, cleaned the extracted text, and parsed each syllabus into sections to isolate course content (see Appendix B.2.1 for details).

Most syllabi follow a common structure: they begin with basic course details (code, title, instructor), followed by a description of the course's content, detailed topic outlines, required readings, evaluation criteria (assignments and exams) and general course policies.

Basic course details We extract course codes, titles, academic terms, years, and instructor names. Using course codes and titles, we categorize each syllabus as lower undergraduate, upper undergraduate, or graduate (see Appendix B.3.2). Texas courses are assigned to a field following the National Center for Education Statistics' Classification of Instructional Programs (CIP). For most analyses, we aggregate fields into four broad macro-fields: STEM, Humanities, Social Sciences, and Business.

Course content We identify the section of a syllabus that provides a description of the course's content by searching for headings such as "Summary," "Description," and "Content" (Appendix Table B2). These sections typically include course structure, main concepts, timelines, and reading materials.

Reference list We extract bibliographic information on required and recommended readings listed on each syllabus (textbooks, articles, and academic papers) using OpenAI's GPT-4.1 mini. We successfully compile reference information for 80.4% of non-empty syllabi; we extract at least one publication year for 42.0% of these syllabi.

[4] We contacted all public universities in Texas that do not make historical syllabi available online to request access to their records. However, most universities were unable to provide these documents because Texas House Bill 2504 of 2009 only requires public colleges and universities to maintain records for two years following the term of instruction.

Sample description Our syllabi and course data are described in panel (a) of Table 1. The average undergraduate syllabus contains 393 unique knowledge terms (i.e., those belonging to a pre-defined dictionary aimed at capturing a document's academic content, defined in greater detail in Section 3), while the average graduate syllabus contains 399 unique terms. Syllabi cover 18,994 undergraduate and 8,878 graduate courses. Most syllabi are from STEM fields; among undergraduate courses, 74% are upper-level (i.e., generally taken during or after the third year).

2.2 Academic Publications

We compiled a database of peer-reviewed articles using OpenAlex, an open-license catalog of global research. Maintained by OurResearch, OpenAlex indexes over 240 million scholarly works, and adds approximately 50,000 new works daily, drawing from multiple sources such as Crossref, PubMed, institutional repositories (e.g., arXiv), and the legacy Microsoft Academic Graph. The platform provides comprehensive metadata on works, authors, institutions, and topics. Our final dataset includes approximately 107 million English-language articles with non-missing information on titles, abstracts, keywords, authors, and author affiliations (see Appendix B.4 for additional details).

2.3 Student Records

We linked syllabi to individual-level administrative records on students enrolled at our seven public universities, provided by the Texas Education Research Center (ERC). Records from the Texas Higher Education Coordinating Board (THECB) contain demographic data (gender and race), family income (from FAFSA), SAT/ACT scores, university enrollment and degree completion records (starting year, type of degree, initially declared major, graduation outcome, graduation year, and full academic transcripts, all available for 2011-2023). Records from the Texas Workforce Commission (TWC) provide quarterly earnings data for all employed individuals in Texas (2011-2025).

Using enrollment records, we assign each student a degree-specific start year, defined as the first year in which the student appears as enrolled in that degree program. We also assign students a major based on the four-digit CIP code declared at enrollment.

Students' academic transcripts list all the courses each student ever enrolled in, regardless of whether they were completed. We use transcripts to link students to the syllabi of their courses, using course codes and years (we ignore section identifiers, since all sections of a course typically share the same syllabus). We successfully link 66% of all courses in the transcript of the average student.

Table 1: Summary Statistics: Syllabi, Instructors, and Students

Panel a) Syllabi and courses				
	Undergraduates		Graduates	
	Mean	SD	Mean	SD
Syllabi				
Unique knowledge words	393.08	284.41	399.31	312.13
Frontier knowledge proximity	105.15	7.57	106.69	7.28
N syllabi	406,969		52,446	
Courses				
Business	0.067	0.251	0.090	0.286
STEM	0.409	0.492	0.428	0.495
Social Science	0.179	0.383	0.325	0.468
Humanities	0.299	0.458	0.134	0.341
Upper level	0.739	0.439		
# sections	4.10	10.93	1.71	2.11
# instructors	2.62	4.45	1.48	1.67
Frontier knowledge proximity	106.04	6.65	107.16	7.24
# courses	18,994		8,878	
N (course * year)	100,746		30,781	
Panel b) Students				
	Undergraduates		Graduates	
Variable	Mean	SD	Mean	SD
Female	0.546	0.498	0.563	0.496
Family income <40K	0.197	0.397	0.200	0.400
Family income >90K	0.202	0.401	0.205	0.404
Field				
Business	0.170	0.375	0.210	0.407
STEM	0.457	0.498	0.298	0.457
Social Sciences	0.208	0.406	0.355	0.478
Humanities	0.158	0.365	0.087	0.282
Courses				
# courses	49.084	32.267	8.965	8.657
Frontier knowledge proximity (in course-level SD)	-0.112	0.514	0.223	0.701
Outcomes				
GPA	2.813	0.783	3.240	0.642
Graduated	0.812	0.391	0.584	0.493
Time-to-degree	5.745	1.278	2.291	1.367
Enrolled in grad school	0.130	0.337		
Quarterly earnings ($1,000)	17.56	20.94	22.93	22.82
N	202,725		41,703	

Note: Summary statistics of the variables used in the analysis.

Given the time span of our syllabi data, we focus on students who started their program between 2011 and 2023. The linked dataset contains 409,672 undergraduate, 73,641 Master's, and 21,493 PhD

students with non-missing syllabi. Our final analysis sample covers 202,725 undergraduate and 41,703 graduate full-time students. We define each student's predicted graduation year as six years after degree start for undergraduates and three years after start for graduate students, and treat quarterly earnings below $1,000 as missing. In our main sample, we retain students with at least one quarter of non-missing earnings during the 1 to 6 years following their predicted graduation year.[5]

Panel (b) of Table 1 describes our student sample. Approximately 55% of undergraduate students and 56% of graduate students are female; about 20% have family income below $40K, and about 20% have family income above $90K. The largest share of undergraduate students (46%) are in STEM fields, while the largest share of graduate students (36%) are in the Social Sciences. On average, undergraduate students enroll in 49 courses, and 81% of them graduate, with a mean time-to-degree of 5.7 years. Thirteen percent enroll in a graduate program at a Texas public university. Graduate students enroll in 9 courses on average and 58% of them graduate, with a mean time-to-degree of 2.3 years. One to six years after predicted graduation, undergraduate students earn $17,560 per quarter; graduate students earn $22,930 per quarter. Additional details on the student data are in Appendix C.

3 Measuring Frontier Knowledge Proximity

We use the text of syllabi and academic publications to construct a course-level measure of frontier knowledge proximity. Our goal is to capture the relative similarity between a syllabus and different vintages of research knowledge: a syllabus has higher proximity to the frontier if the concepts, methods, applications, and evidence it emphasizes are more similar to recent research than to older research. We now detail the steps to construct this measure and present a series of checks to validate it. Additional details on the construction of the measure are provided in Appendix B.5.

3.1 Similarity Between Syllabi and Academic Publications

Our procedure largely follows Biasi and Ma (2022). We start by representing each document d (a syllabus or an article) as a term-frequency vector $\mathbf{TF_d}$ of length $|W|$, where W is a list of all the

[5]Our results on educational outcomes remain robust if we include all students for whom we have syllabi information, regardless of whether we observe their earnings; see Appendix Table A5. Appendix Table A4 also shows that frontier exposure does not significantly predict having no observed earnings or having non-missing earnings for at least three consecutive quarters.

terms we consider. Each element TF_{dw} of $\mathbf{TF_d}$ is the frequency of term w in d:

$$TF_{dw} \equiv \frac{c_{dw}}{\sum_{k \in W} c_{dk}},$$

where c_{dw} is the number of times term w appears in d and the denominator is the total number of terms in d. To focus on academic knowledge, we construct W as the list of all unique terms used as keywords in our academic publications sample.[6]

Adjusting for Term Relevance A standard issue with term frequency vectors is that terms commonly used across *all* documents mechanically receive higher weight, regardless of their ability to capture the content of each specific document. For example, the term "Programming" appears frequently in Computer Science syllabi but is less informative than "Natural Language Processing." Similarly, "Animals" is common in Biology syllabi but less informative than "CRISPR" (a gene-editing technology).[7] The text analysis literature addresses this issue with term-frequency-inverse-document-frequency (TFIDF) weighting (Kelly et al., 2021), implemented by multiplying each element in a word vector by the inverse of the term's frequency across all documents:

$$TFIDF_{dw} \equiv TF_{dw} \times IDF_w, \tag{1}$$

where IDF_w is the inverse ratio of the count of all documents containing word w, $\sum_{n \in D} \mathbb{1}(c_{nw} > 0)$, and the count of all documents, $|D|$:

$$IDF_w \equiv \ln\left(\frac{|D|}{\sum_{n \in D} \mathbb{1}(c_{nw} > 0)}\right).$$

Accounting for Changes in Term Relevance Over Time Since our goal is to capture content that is closer to the current research frontier, an unappealing feature of traditional TFIDF weights is that they ignore temporal changes in term relevance. Terms that became popular only recently receive low weights even in older documents due to their current widespread use. Consider course CS229 at Stanford University, taught by Andrew Ng in the early 2000s as one of the first courses entirely focused on machine learning. The term "machine learning" has become very popular in later years, so its frequency across all documents is high and its IDF_w is low. Pooling documents from different years would thus assign a low $TFIDF_{dw}$ to "machine learning" in this course's syllabus, failing to

[6] Our results are robust to considering all terms with an English Wikipedia webpage as of 2019 (results are available upon request).

[7] CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats, a family of DNA sequences found in prokaryotic organisms and the basis for gene-editing technology.

recognize its novelty in the early 2000s.

We address this issue by adjusting the TFIDF weighting scheme to incorporate a backward-looking IDF measure, $BIDF_{tw}$, based only on documents published prior to year t:

$$BIDF_{tw} \equiv \ln\left(\frac{|D_t|}{\sum_{n\in D_t}\mathbb{1}(c_{nw}>0)}\right),$$

where D_t is the set of documents published *prior to* t. Using $BIDF_{tw}$, we construct a term-frequency-backward-inverse-document-frequency vector $\mathbf{TFBIDF_d}$, with elements:

$$TFBIDF_{dw} = TF_{dw} \times BIDF_{t(d)w}, \tag{2}$$

where $t(d)$ denotes the publication year of document d. In our example, "machine learning" would have a higher $BIDF$ in the early 2000s than in 2018 (when the term was widely used), appropriately assigning a higher $TFBIDF$ to course CS229 and better capturing its novelty.

Textual Similarity Between Documents Equipped with weighted vectors $\mathbf{TFBIDF_d}$, we calculate the textual similarity between pairs of documents d and d' as the cosine similarity between their weighted word vectors (for simplicity, we denote $\mathbf{TFBIDF_d}$ as $\mathbf{V_d}$):

$$\rho_{d,d'} = \frac{\mathbf{V_d}}{\|\mathbf{V_d}\|}\cdot\frac{\mathbf{V_{d'}}}{\|\mathbf{V_{d'}}\|} \tag{3}$$

where $\|\mathbf{V_d}\|$ is the Euclidean norm of $\mathbf{V_d}$. Since each element of $\mathbf{V_d}$ is non-negative, ρ lies in $[0,1]$. If d and d' are from the same year and use the exact same set of terms with the same frequency, $\rho_{d,d'}=1$; if they have no terms in common, $\rho_{d,d'}=0$.

3.2 Calculating Frontier Knowledge Proximity

To capture the similarity between each syllabus d and different vintages of research, we calculate the average similarity of d with all articles published in a three-year window centered τ years before the syllabus year $t(d)$:

$$S_d^\tau = \frac{\sum_{n\in\Omega_\tau(d)}\rho_{dn}}{|\Omega_\tau(d)|}$$

where $\Omega_\tau(d)$ includes all articles published in the period $[t(d)-\tau-1, t(d)-\tau+1]$, and $|\Omega_\tau(d)|$ is the number of these articles.[8]

Our frontier knowledge proximity measure is the ratio between the average syllabus similarity

[8]Our main analysis uses three-year intervals; results are robust to one-year or two-year intervals.

with recent articles (published in $[t(d) - \tau - 1, t(d) - \tau + 1]$) and older articles (published in $[t(d) - \tau' - 1, t(d) - \tau' + 1]$, with $\tau' > \tau$), multiplied by 100 for readability:

$$p_d \equiv 100 \times \left(\frac{S_d^{\tau}}{S_d^{\tau'}} \right). \tag{4}$$

A syllabus taught in year t thus has higher proximity if it is more similar to recent research than to older research. This measure is the inverse of the *education-innovation gap*, which we first proposed in Biasi and Ma (2022).

In setting the parameters τ and τ', which define recent and older knowledge vintages, we account for field-specific rates of knowledge production. In fields where knowledge production is fast, both τ and τ' are small; in fields where it is slower, τ' (and possibly τ) are large. We define τ and τ' as the 5th and 90th percentiles of the distribution of citation ages within each field across all articles in that field. The median recent knowledge vintage (5th percentile) is approximately 2 years, while older vintages (90th percentile) have a median of 16 years, ranging from 12 years in fast-moving fields like Medicine to 36 years in slower-moving fields such as Religion (Appendix Figure A1).

Our measure has two appealing properties. First, as a ratio, it is less sensitive to stylistic differences across syllabi (such as length or verbosity) that could otherwise affect similarity scores. Two courses covering the same materials could have different similarities to research publications if one syllabus is more detailed or uses more academic terms. Using a ratio attenuates such variations since both the numerator and denominator are subject to the same style differences. Second, by employing the $TFBIDF$ weighting scheme, the measure appropriately reduces the influence of common foundational terms. This ensures that syllabi covering widely taught, fundamental, but old concepts (i.e., the "classics" of a field, such as *Ordinary Least Squares* in applied microeconomics) are not mechanically favored or penalized by raw term prevalence.

3.3 Validating Frontier Knowledge Proximity

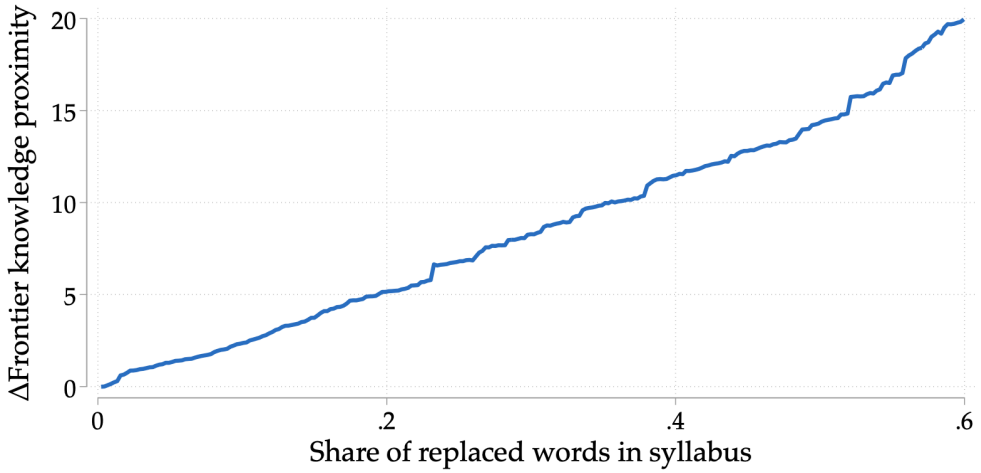
We validate whether the measure captures how close course content is to the research frontier using four complementary tests.

AI-based content validation Our first test, described in detail in Appendix B.6, probes the substantive content captured by the measure. This exercise asks whether a model-based assessment of real syllabi ranks courses in the same direction as frontier knowledge proximity. For a stratified sample of 50 pairs of comparable courses, we use an LLM to extract the ten main subject-matter

knowledge points from each syllabus, excluding course logistics, grading rules, teaching strategies, and bibliographic metadata. We then ask the model to compare the two lists and identify which syllabus covers more frontier-oriented concepts, methods, or tools. In 37 of 50 pairs (74%), the model's majority judgment agrees with the ranking implied by our measure. We view this exercise as reassuring evidence that the measure captures substantive differences in frontier-oriented course content.

Simulation exercise We additionally use a simulation to demonstrate that our measure can detect incremental changes in a syllabus's coverage of knowledge vintages. In a random sample of 100,000 syllabi, we progressively replace terms that appear more frequently in older articles ("old terms") with terms that appear more frequently in recent articles ("new terms"), then recalculate proximity at each incremental replacement. To ensure a monotonic increase, we use a recursive structure: to replace n terms, we sample from the top $2n$ terms with the highest "oldness" scores, ensuring the set of n terms contains all terms from the $n-1$ step. We define old terms as either: (a) within the top 5% frequency among older articles (published between $t-\tau'-1$ and $t-\tau'+1$), or (b) appearing in the old corpus but not in the recent corpus (published between $t-\tau-1$ and $t-\tau+1$), where t denotes the syllabus year, and τ and τ' are field-specific as defined in Section 3.2. We define new terms symmetrically as either (a) in the top 5% in terms of frequency in the new publication corpus, or (b) in the new corpus but not in the old corpus.

Figure 1: Updates in Syllabi Content and Changes in Frontier Knowledge Proximity



Note: This figure illustrates the median change in proximity when we manually replace "old" knowledge words with "new" knowledge words in a random sample of 100,000 syllabi.

Proximity increases monotonically as we replace more old terms with new ones. Figure 1 plots

the median change in proximity across all syllabi for a given number of replaced terms. A one-SD increase in frontier knowledge proximity (approximately 7.5) is equivalent to replacing 27% of the old content in the median syllabus with new content.

Reference age We verify that frontier proximity correlates negatively with the average age of a syllabus's references, calculated as the average difference between the syllabus year and the publication years of cited sources (Figure A2, panel (a)). The correlation is modest in magnitude (-0.01). This weak relationship is unsurprising, as many syllabi primarily reference textbooks whose publication dates might not accurately reflect course content. While reference age is easy to compute, our text-based measure better captures actual course content and is available even for syllabi that lack explicit references or rely primarily on a single textbook.

Differences by course level Finally, we confirm that proximity varies as expected across course levels. Graduate courses and upper undergraduate courses have higher proximity scores than lower undergraduate courses (Figure A2, panel (b)), consistent with the expectation that more advanced courses incorporate more frontier knowledge.

3.4 Differences in Frontier Knowledge Proximity Across Courses

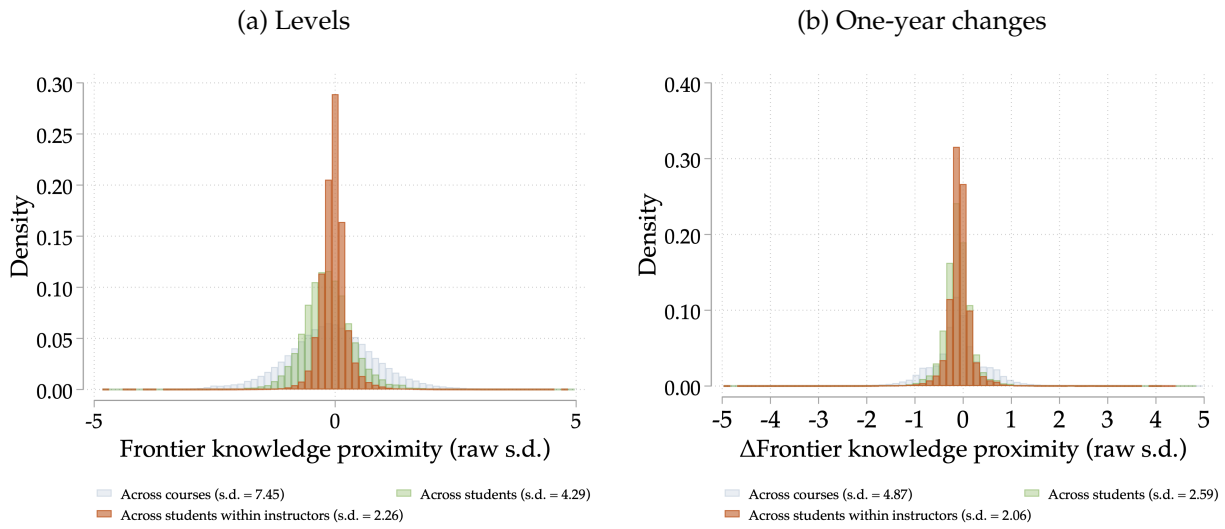
We begin our analysis by examining the variation in frontier knowledge proximity across courses in our samples. On average, this measure equals 105.2, with a standard deviation of 7.45, a 25th percentile of 101.2, and a 75th percentile of 109.1. Figure 2 (panel a, lighter series) shows the distribution of frontier knowledge proximity across courses, normalized so that the mean is zero.

To give these numbers an economic interpretation, we rely on the simulation in Figure 1. A one–SD increase in proximity corresponds to replacing 27% of the knowledge content of the median Texas syllabus. Moving a Texas syllabus from the 25th to the 75th percentile of proximity (a 7.9 increase) corresponds to replacing approximately 30% of the median syllabus content. These findings illustrate substantial variation in frontier knowledge proximity across courses.

3.4.1 Explaining the Variation in Frontier Knowledge Proximity

Several factors could explain this variation. Fields may differ in their emphasis on foundational versus frontier knowledge; for example, mathematics courses might emphasize more foundational content compared to computer science courses. Institutional characteristics, such as teaching philosophy, resources, or faculty expertise, might also matter; for instance, less-resourced institutions may face challenges integrating cutting-edge material. Differences between instructors, as well as variations over time, might further influence proximity.

Figure 2: Distribution of Frontier Knowledge Proximity



Note: Distribution of frontier knowledge proximity measures, in levels (panel a) and in changes from the previous year (panel b).

To quantify the contribution of each of these factors to explaining the variation, we perform a Shapley-Owen decomposition of variance (Israeli, 2007; Huettner et al., 2012). Specifically, we estimate the adjusted R^2 of a regression of proximity on fixed effects for fields, schools, years, instructors, and courses. We then compute each factor's partial contribution (partial-R^2) by measuring the average decline in adjusted R^2 when removing that factor from the regression.[9]

The results of this decomposition exercise, shown in Appendix Table A1, indicate that differences across instructors and courses explain the largest portion of the overall variation in frontier knowledge proximity—21% and 26%, respectively. By contrast, field effects explain only 5%, and school effects contribute just 0.3%.

4 The Effects of Frontier Knowledge on Students: Research Design

Our descriptive analysis documents significant differences in frontier knowledge proximity across courses. Do these differences matter for students' outcomes? We begin by developing an empirical strategy to isolate the causal effect of frontier knowledge exposure when students can select courses.

[9]The decomposition involves three steps: (i) estimate OLS models of proximity on all possible combinations of factors and record their adjusted R^2; (ii) re-estimate these models excluding one factor at a time and record the decrease in adjusted R^2; (iii) calculate the average decline in adjusted R^2 for each factor across all possible combinations, defining this average as the partial-R^2 attributable to each factor. For factor j, the partial-R^2 is calculated as $R^2_j = \sum_{T \subseteq V \setminus \{j\}} \frac{|T|!(K-|T|-1)!}{K!}[R^2(T \cup \{j\}) - R^2(T)]$, where $R^2(S)$ is the adjusted R^2 of a regression of proximity on the set of factors S, V is the full set of factors, $|T|$ denotes the number of factors in subset T, and $K \equiv |V| = 5$ is the total number of factors. We use the adjusted R^2 to ensure comparability across factors with different numbers of categories.

The next section presents our results.

4.1 Empirical Model

The core empirical model we want to estimate is

$$y_i = \beta p_i + \xi X_i + \theta_{s(i)m(i)c(i)} + \varepsilon_i, \tag{5}$$

where y_i denotes student i's outcome (e.g., average earnings 1-6 years after predicted graduation) and p_i is the average frontier knowledge proximity of all courses on student i's transcript. The vector X_i includes individual characteristics that may affect outcomes, such as gender, race, family income, and prior academic achievement. The term θ_{smc} denotes school s-by-major m-by-starting cohort c fixed effects.

Our parameter of interest is β, which captures the causal effect of average course proximity p_i on student outcomes within groups of students who enter the same program (i.e., school-major) in the same year. The main challenge in estimating β is that p_i is a function of students' course choices, and therefore may be correlated with unobserved characteristics that also affect outcomes. For example, higher ability students may select courses with higher frontier knowledge proximity and earn more after graduation. This implies that $\mathbb{E}(p_i \varepsilon_i) \neq 0$ and leads to bias in OLS estimates of β.

4.2 Identification Strategy

To overcome this challenge, we exploit an institutional feature that generates variation in p_i that is unobservable to students at the time of course selection. Under Texas House Bill 2504 (enacted in 2009), instructors at public universities are required to post syllabi online only *after* the course has started, not before registration. Course registration at Texas universities takes place several weeks before the semester begins, so when students make enrollment decisions, they are not guaranteed access to the updated syllabus for the upcoming term. In a systematic audit conducted in 2025, we verified that at every institution in our sample, syllabi were generally not visible on the university's course registration portal at the time of registration (Appendix Table A2). We also confirmed each university's stated syllabus-posting rule one by one on institutional websites (Appendix D).[10] The

[10] Appendix D documents in detail the plausibility of the identification assumption at each institution in our sample. We find strong institutional evidence for all institutions except for the University of Houston Clear Lake. Despite the validity of House Bill 2504, the university website does not itself state that the public posting deadline occurs after the first class day. Our results remain robust when we exclude this university from the estimation sample (Appendix Tables D1 and D2).

information available to students at registration is therefore the syllabus from the prior term and the name of the instructor. Consequently, conditional on the instructor and the course's frontier knowledge proximity in the previous term, the current-term proximity of a given course is unobserved by students at the time of enrollment and therefore quasi-random from their perspective.

To formalize this argument, consider a student i choosing between two courses, A and B, during term t. Let $p_{A,t}$ and $p_{B,t}$ denote their frontier knowledge proximity and γ_A and γ_B be indicators for their instructors. The student's realized proximity, p_i, equals $p_{A,t}$ if the student enrolls in course A and $p_{B,t}$ if she enrolls in course B. Define the change in proximity for course k as $\Delta_{k,t} \equiv p_{k,t} - p_{k,t-1}$ for all $k \in \{A, B\}$. We can express the realized proximity as

$$p_i = g\Big(\overbrace{\alpha_i, \nu_i, \underbrace{p_{A,t-1}, p_{B,t-1}, \gamma_A, \gamma_B}_{\text{course content}}}^{\text{course selection}}, \Delta_{A,t}, \Delta_{B,t}\Big),$$

where α_i represents student i's unobserved preference for high-proximity courses and ν_i includes other unobserved factors that may affect course selection. In words, students' course choices depend only on α_i, ν_i, $p_{A,t-1}$, $p_{B,t-1}$, γ_A, and γ_B, whereas course content is determined by $p_{A,t-1}$, $p_{B,t-1}$, γ_A, γ_B, Δ_A, and Δ_B. Endogeneity arises if ν_i and α_i are correlated with the unobserved component of the outcome (denoted by ε_i in equation (5)).

Our identification argument relies on the exogeneity of the underlying course updates, $\Delta_{A,t}$ and $\Delta_{B,t}$, with respect to the student unobservables α_i, ν_i, and ε_i. Because these updates are unobserved at registration, they cannot drive a student's course choice. Therefore, conditional on $p_{A,t-1}$, $p_{B,t-1}, \gamma_A$, and γ_B—the observable variables driving course selection—the remaining variation in actual proximity p_i is driven entirely by these unexpected updates, rendering it conditionally exogenous. In this simple example with two courses, we can therefore consistently estimate β by augmenting the model in equation (5) with controls for these variables:

$$y_i = \beta p_i + \xi X_i + \delta_A p_{A,t-1} + \delta_B p_{B,t-1} + \theta_{s(i)m(i)c(i)} + \gamma_A + \gamma_B + \varepsilon_i.$$

4.2.1 Implementation

In practice, students choose multiple courses from a large set of potential courses; the average undergraduate student in our data reports 49 courses on her transcript (Table 1). Our identification argument relies on the conditional independence of p_i and ε_i given the lagged proximity of all courses in i's choice set $\mathcal{C}_i$ and the identity of i's instructors. Assuming that all students in the same

program and starting cohort face the same choice set, we can approximate $\mathcal{C}_i$ with the set of all courses ever taken by students in the same school, major, and cohort. Given the size of $\mathcal{C}_i$, directly including the lagged proximity of all courses in a student's choice set is infeasible. Instead, we summarize it with two aggregate measures: $p_{i,-1}$, the average previous-year proximity of courses that the student takes, and p_{-1}, the average previous-year proximity of all courses in the student's choice set, which in our empirical model is subsumed by the school-by-major-by-cohort fixed effects.

We therefore implement our identification strategy by augmenting equation (5) with controls for $p_{i,-1}$ and fixed effects for the instructors of i's courses. The estimating equation becomes:

$$y_i = \beta p_i + \delta p_{i,-1} + \xi X_i + \sum_{k \in \mathcal{K}_i} \gamma_k + \theta_{s(i)m(i)c(i)} + \varepsilon_i, \tag{6}$$

where γ_k is a fixed effect for the instructor of course k and $\mathcal{K}_i$ is the set of all instructors who ever taught student i. We use robust standard errors unless stated otherwise.[11]

Our framework assumes that the decision to take course k at time t is made independently for each course and term. Because students often face structured prerequisites and required course sequences, this is a strong simplification. However, by proxying the program-level menu with fixed effects and controlling for the student's realized lagged exposure, the remaining variation isolates unanticipated syllabus updates within a given instructor's course. Our estimand is therefore the local effect of experiencing a surprise instructor-led modernization of course content.

Identifying variation Our research design uses variation in course proximity among students who enrolled in the same program in the same year *and* were taught by the same instructors. Since instructors explain a significant portion of the total variation in proximity across courses (Table A1), one might ask whether we have enough variation left to identify causal effects. Figure 2 contrasts the distribution of our raw proximity measure across all courses (in blue), across all students (thereby giving more weight to courses with larger classes, in green), and across students within school-major-cohort and instructor cells. The standard deviation equals 7.45 across courses, 4.29 across all students, and 2.26 across students within cells (panel (a)). The standard deviation of the proximity change across students within instructors is 2.06 (panel (b)). The simulation exercise in Figure 1 implies that the latter corresponds to a 10% change in the content of the average course taken by each student.

[11] An alternative way to compute standard errors in our context would be to cluster them at the program-cohort level. This procedure generally yields smaller standard errors, so we view our current approach as conservative. These results are available upon request.

4.2.2 Identifying Assumptions

Our identification strategy rests on three main assumptions.

The first assumption is that $\Delta_{k,t}$ must be exogenous. This implies that students cannot predict syllabus updates when choosing courses. Even if students cannot see updated course syllabi at the time of enrollment, this assumption could be violated if they can predict content changes via observable changes in other course attributes. The most obvious is a change in instructor. To address this issue, we explicitly control for instructors in our main equation. We additionally control for changes in course content that are common across courses in a given program (e.g., those driven by the directives of a new department chair) by including school-by-major-by-cohort fixed effects in all our specifications.

This assumption could also be violated due to the add/drop window at the start of each term. While registration takes place before the semester begins, students may add courses for a few days after the term starts—at which point the updated syllabus has been posted—and drop courses for a somewhat longer window. Both actions introduce selection on $\Delta_{k,t}$. We address each concern separately. For course drops, we assign each course to a student based on whether they ever enrolled in it, regardless of whether the course was subsequently completed; our estimates of β are consequently intent-to-treat (ITT) effects. In our data, 28% of students drop at least one course in a given year and each student drops one course per year on average. For late adds, we cannot observe registration dates and therefore cannot directly identify late adders. However, since adding a course late generally requires first dropping another, students who drop any course in a given semester are the most likely to include late adders. Moreover, many Texas public universities explicitly treat such requests as exceptional or disfavored: adding a course after the ordinary registration/add deadline requires additional approvals and often triggers late-registration or late-add fees.[12] As a robustness check, we re-estimate our main specifications excluding students who drop any course in a given year; results are unchanged (Appendix Tables A12 and A13).

The second assumption is that course design is independent of student characteristics, requiring that instructors design their courses without regard to the characteristics of future students. This assumption is plausible because syllabi are generally compiled before instructors meet their students. In support of it, in Appendix Table A3 we regress absolute course proximity ($p_{k,t}$, column 1) and the

[12]For example, at the University of Texas at Dallas, adding a course after Census Day requires instructor approval and is subject to a per-course late add fee (currently $100) (https://catalog.utdallas.edu/2023/undergraduate/policies/registration-and-enrollment); similarly, at Texas A&M University, adding courses after the first day of classes requires additional approvals and incurs late registration and late add fees (https://aggie.tamu.edu/registration-and-records/classes/registration-information/late-registration).

change in course proximity from the previous term ($\Delta_{k,t}$, column 3) on detailed characteristics of all students taking the course. While these characteristics appear correlated with $p_{k,t}$ (a clear indicator of endogenous course selection, column 1) they do not predict $\Delta_{k,t}$: the p-value of an F-test of joint significance is 0.67 (column 3).

The third and perhaps most important assumption, essential for interpreting our estimates as the effects of frontier knowledge exposure, is that changes in a course's proximity do not occur at the same time as changes in other dimensions of content or instruction. This assumption could be violated if, when updating frontier content, instructors also change course design (for example, shifting evaluation from exams to reports or presentations) in ways that directly affect outcomes. To assess these possibilities, we regress $p_{k,t}$ and $\Delta_{k,t}$ on syllabus and course characteristics (average grades, an indicator for whether the course requires an exam, and the shares of total assessment assigned to homework and reports) and instructor characteristics (the total number of instructors, whether any instructor is new to the course, and average instructor experience). Syllabus characteristics do not jointly predict proximity changes (F-statistic p-value = 0.13; Appendix Table A3, column 4) and our estimates are robust to the inclusion of controls for syllabus grading structure (shares of total assessment allocated to exams, homework, and reports; Appendix Tables A14 and A15).

The assumption would also fail if course updates occur at specific points of instructors' careers when effort or teaching effectiveness are also changing. Column 4 of Appendix Table A3 shows that instructor characteristics marginally predict proximity changes. Our baseline specifications control for instructor fixed effects, which absorb time-invariant instructor heterogeneity. In robustness tests, we show that our results remain stable when we control for instructor experience (Appendix Tables A16 and A17) and teaching load, a proxy for time available for course preparation (Appendix Tables A18 and A19).

5 Frontier Knowledge and Students' Outcomes

Using this empirical strategy, we now study the effects of exposure to frontier knowledge on students' academic and labor-market outcomes. We focus on degree completion, time-to-degree, academic performance, graduate-school enrollment, and post-college earnings.

The effects of frontier knowledge on these outcomes are ambiguous ex ante. On the one hand, frontier knowledge may increase the perceived returns to effort and strengthen incentives to persist in college by making coursework appear more relevant to current research and real-world applica-

tions. It may also shorten time-to-degree by revealing the structure, demands, and potential payoffs of advanced coursework earlier in students' college careers, allowing them to better align effort and course choices. Lastly, exposure to frontier knowledge may directly enhance skill acquisition by deepening conceptual understanding, strengthening problem-solving abilities, and improving the capacity to apply knowledge to new contexts. On the other hand, frontier knowledge may negatively affect students by crowding out useful foundational knowledge or by making material too difficult.

To shed light on each of these mechanisms, we estimate not only whether frontier exposure helps, but which margin it shifts. If frontier exposure mainly affects whether students stay engaged and persist, its effects should appear early in college and on fragile completion margins. If it also builds human capital with labor-market value, its returns should accumulate across the curriculum, show up in later coursework, and persist into earnings. The empirical patterns we document are consistent with frontier knowledge improving outcomes heavily through a persistence/engagement-type channel, with a human-capital/skill-building channel emerging for specific subgroups.

5.1 Frontier Knowledge Improves Academic Outcomes

5.1.1 Undergraduate Students

For the average undergraduate student, exposure to frontier knowledge in college has positive effects on all of our academic outcomes. Panel (a) of Table 2 shows estimates of β in equation (6), with various educational outcomes as the dependent variables. A one-SD increase in average course proximity raises the probability of completing a bachelor's degree by 4.8 pp, or 5.5% relative to a mean of 87.3% (Table 2, panel a, column 1). It also helps students graduate faster by reducing time-to-degree by 0.21 years, or 3.8% relative to a mean of 5.67 years (column 2). In addition, frontier knowledge improves students' performance during college: a one-SD increase in proximity raises GPA by 0.12 grade points (column 3).

A possible concern with these results is that, because frontier knowledge tends to be higher for more advanced courses, it may mechanically be higher for students who progress toward the latest stages of their degree, creating a spurious relationship with outcomes such as completion and time-to-degree. However, our identification strategy isolates the effects of changes in frontier knowledge (rather than levels) on outcomes. Furthermore, our estimates remain qualitatively unchanged if we restrict attention to the subsample of students who reach year 3 of their program: the estimated effects are 4.2 pp on degree completion, -0.21 years on time-to-degree, and 0.08 GPA points (Appendix

Table A6).

Exposure to frontier knowledge in college also makes students significantly more likely to attend graduate school. A one-SD increase in proximity raises the probability of graduate-school attendance by 1.4 pp, or about 12% relative to a mean rate of 11.6% (Table 2, panel a, column 4). Taken together, these estimates show that frontier exposure improves undergraduate attainment on every measured margin.[13]

5.1.2 Graduate Students

Frontier knowledge also benefits graduate students, with effects concentrated among Master's students. For Master's students, a one-SD increase in proximity raises graduation rates by 1.6 pp (about 2.2% relative to a mean of 71.4%; Table 2, panel (b), column 1), reduces time-to-degree by 0.12 years (column 2), and raises GPA by 0.03 points (column 3), though only the time-to-degree estimate is statistically distinguishable from zero. For PhD students, the same increase in proximity does not have a detectable effect on graduation, time-to-degree, or GPA (columns 4-6). These results indicate that frontier content is not just useful for undergraduates on the margin of finishing college; it can also speed progression in advanced training, especially in Master's programs.

5.2 Exposure to Frontier Knowledge Increases Post-College Earnings

Exposure to frontier knowledge leads to significant increases in earnings for both undergraduate and graduate students. Estimates of equation (6), using the log of average quarterly earnings 1-6 years after predicted graduation (defined as 6 years after degree start for undergraduates and 3 years after start for graduates), indicate that a one-SD increase in frontier knowledge proximity raises earnings by about 2.1% for undergraduates (Table 3, panel (a), column 1) and 4.5% for graduates (panel (b), column 1).

In columns 2-3 of Table 3, we examine the dynamics of these earnings effects by estimating impacts separately for earnings 1-3 years and 4-6 years after predicted graduation. For undergraduates, the estimated earnings gains are larger in the first three years after predicted graduation: a one-SD increase in proximity raises earnings by 3.1% after 1-3 years, while the estimate for years 4-6 is smaller, at 1.3%, and statistically indistinguishable from zero (Table 3, panel (a), columns 2-3).[14] Graduate students display a similar pattern, with a 7.4% increase after 1-3 years and a smaller, sta-

[13] Appendix Figure A3 shows that the main patterns are not driven by a single macro-field.

[14] Since frontier knowledge exposure raises the likelihood of attending graduate school (Table 2, column 4), these short-run earnings estimates may understate the early-career labor-market gains for treated students who spend part of this period enrolled in graduate programs.

Table 2: Educational Effects of Frontier Knowledge Exposure

Panel (a) Undergraduate students					
	Graduates	Time-to degree	GPA	Attends grad school	Future grade rank
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.048***	-0.213***	0.120***	0.014***	0.004***
	(0.005)	(0.023)	(0.013)	(0.005)	(0.001)
Mean dep. var.	0.873	5.668	2.885	0.116	0.524
R^2	0.563	0.596	0.450	0.286	0.085
N	190,439	155,030	189,431	190,439	3,673,834

Panel (b) Graduate students						
	MA			PhD		
	Graduates	Time-to degree	GPA	Graduates	Time-to degree	GPA
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.016	-0.119***	0.025	0.006	0.068	0.033
	(0.012)	(0.038)	(0.017)	(0.019)	(0.393)	(0.044)
Mean dep. var.	0.714	2.295	3.355	0.258	6.300	2.810
R^2	0.509	0.695	0.516	0.818	0.890	0.785
N	26,440	21,949	26,244	4,832	1,248	4,811

Notes: OLS estimates. Panel (a) reports results for undergraduate students and panel (b) for graduate students. The dependent variable is degree completion (column 1 in panel (a) and columns 1 and 4 in panel (b)), time-to-degree in years (column 2 in panel (a) and columns 2 and 5 in panel (b)), GPA (column 3 in panel (a) and columns 3 and 6 in panel (b)), graduate school attendance within Texas public universities (column 4 in panel (a)), and the student's rank in future course grades (column 5 in panel (a)). In all columns except column 5 of panel (a), one observation is a student; in column 5 of panel (a), one observation is a student-course pair. *Proximity* is the average frontier knowledge proximity of the courses taken by each student, measured in course-level standard deviations; in column 5 of panel (a), it is calculated using only courses taken before the focal course. All specifications control for average lagged proximity, instructor fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses in all columns except column 5 of panel (a), where standard errors are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

tistically insignificant 3.2% estimate after 4-6 years (panel (b), columns 2-3). This early concentration of earnings gains is consistent with improved short-run labor-market placement.

Distributional Effects of Frontier Knowledge Exposure We next examine how frontier knowledge proximity affects the likelihood of being in different parts of the earnings distribution. We do so by re-estimating equation (6) using, as outcomes, indicators of having earnings in the top and bottom quartiles of the cohort-specific earnings distribution.

Exposure to frontier knowledge significantly reduces the likelihood of low earnings for under-

Table 3: Earnings Effects of Frontier Knowledge Exposure

Panel (a) Undergraduate students						
	ln(earnings)			Prob. earnings in...		ln(industry
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile	earnings)
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.021**	0.031**	0.013	-0.021***	0.006	0.003
	(0.011)	(0.012)	(0.016)	(0.007)	(0.005)	(0.007)
R^2	0.297	0.302	0.357	0.216	0.316	0.277
N	189,937	157,723	85,958	189,937	189,937	189,937
Panel (b) Graduate students						
	ln(earnings)			Prob. earnings in...		ln(industry
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile	earnings)
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.045**	0.074***	0.032	-0.015	0.019*	0.013
	(0.018)	(0.022)	(0.033)	(0.009)	(0.011)	(0.009)
R^2	0.525	0.553	0.571	0.467	0.525	0.568
N	30,793	25,201	13,574	30,793	30,793	30,793

Notes: OLS estimates; one observation is a student. The dependent variable is the natural logarithm of earnings 1-6 years (column 1), 1-3 years (column 2), and 4-6 years after predicted graduation (column 3); an indicator for earnings in the bottom quartile (column 4) and the top quartile of the graduation cohort-specific earnings distribution (column 5); and the natural logarithm of mean earnings in the first industry of employment post-graduation (column 6). Panel (a) shows estimates for undergraduate students and panel (b) shows estimates for graduate students. The variable *Proximity* is the average frontier knowledge proximity of all courses taken by each student, measured in course-level standard deviations. All specifications control for average lagged proximity, instructor and school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

graduates and raises the likelihood of high earnings for graduates. A one-SD increase in frontier exposure during undergraduate studies lowers the likelihood of being in the bottom earnings quartile by 2.1 pp (Table 3, panel (a), column 4), while the estimate for the top quartile is small and statistically indistinguishable from zero (column 5). The same increase during graduate studies lowers the probability of bottom-quartile earnings by 1.5 pp and raises the probability of top-quartile earnings by 1.9 pp (Table 3, panel (b), columns 4 and 5). This pattern suggests that undergraduate exposure mainly reduces downside risk, whereas graduate exposure increases the likelihood of reaching the top of the earnings distribution.

The Role of Industry Sorting A possible channel through which frontier knowledge may increase earnings is by raising the likelihood of employment in high-paying industries. To test for this, we calculate the average earnings in each student's industry of first employment and use this measure as the outcome in equation (6). Estimates are small and statistically indistinguishable from zero in both the undergraduate and graduate samples (Table 3, column 6). These results suggest that the earnings gains are not primarily driven by sorting into higher-paying industries.

5.3 Dynamic Evidence: Frontier Knowledge and Future Grades

The earnings effects are consistent with a model where frontier knowledge promotes learning and skill development. Positive effects of frontier exposure on GPA support this hypothesis. Yet they could also accrue if courses with higher proximity are graded more leniently.

To better isolate the effects on learning, we exploit the time dimension of students' transcripts and study whether exposure to frontier knowledge in earlier courses affects performance in subsequent courses. We modify our estimating equation as follows:

$$y_{i,k} = \tilde{\beta} p_{i,t(k)} + \delta p_{i,-1,t(k)} + \xi X_i + \gamma_{k(i)} + \theta_{s(i)m(i)c(i)} + \varepsilon_{ik}, \tag{7}$$

where $y_{i,k}$ is student i's grade rank in course k, $t(i,k)$ is the year in which i takes course k, $p_{i,t}$ is the average proximity of all courses taken by student i prior to t, and $p_{i,-1,t}$ is its lagged counterpart. Everything else is as before.

Estimates for undergraduate students indicate that a one-SD increase in the average frontier knowledge proximity of prior courses raises the student's rank in future courses by 0.4 percentile points (Table 2, panel (a), column 5, significant at 1%). Although this estimate is modest in magnitude, it is difficult to reconcile with a pure grading-leniency story and is more consistent with cumulative learning or skill accumulation from frontier knowledge exposure.

5.4 When in the Curriculum Does Frontier Knowledge Exposure Matter?

We next examine whether the timing of frontier knowledge exposure during a student's college career matters for impacts. If frontier knowledge affects outcomes primarily by enhancing persistence and motivation, the effect should be largest in earlier years, when students are still deciding whether to stay on track. If it operates primarily through skill accumulation, the effect may emerge later, when coursework is more specialized and closer to labor-market entry.

To test whether the effect of frontier knowledge proximity differs across the various stages of the college journey, we focus on undergraduate students and estimate a student-year stacked panel

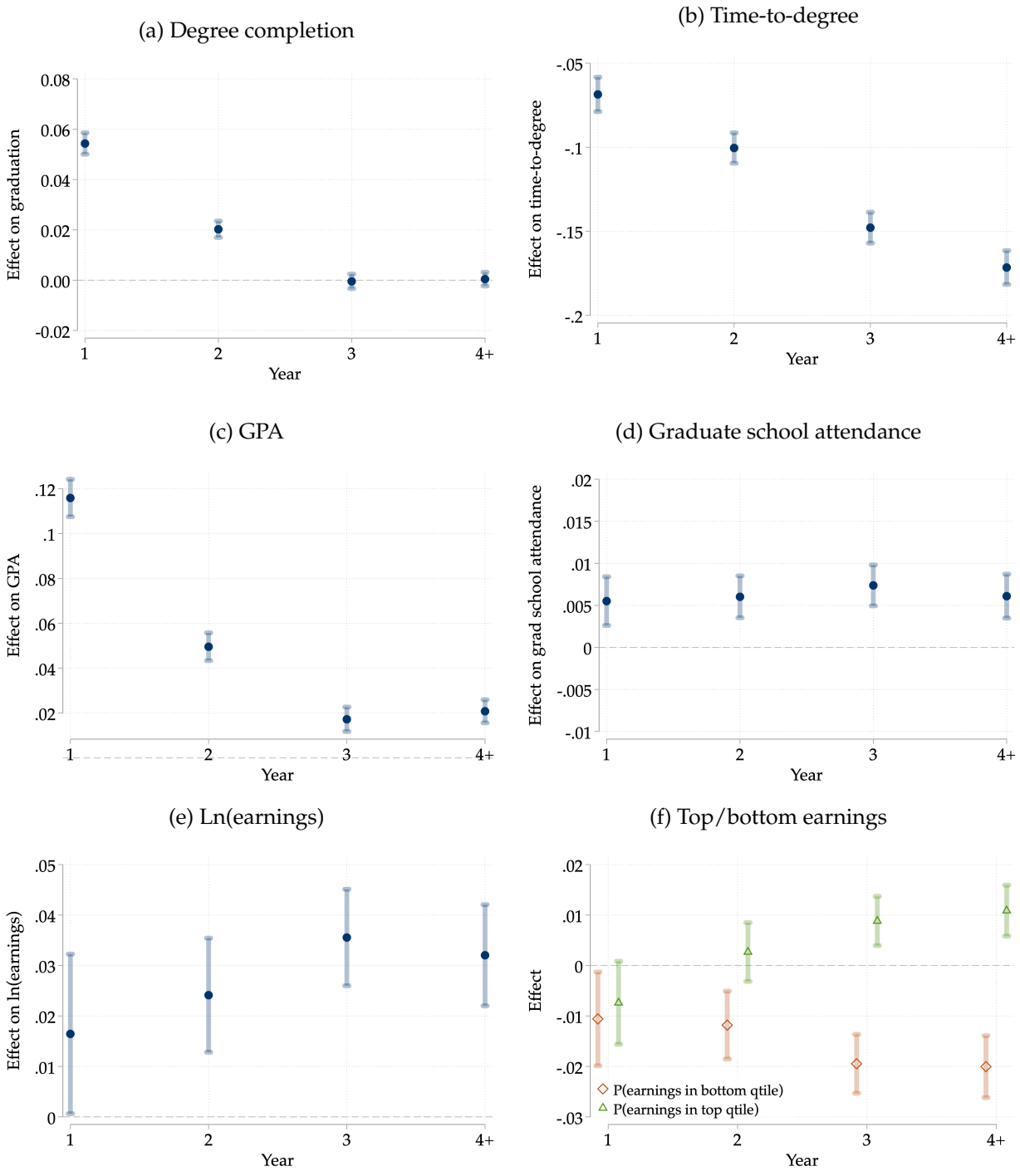
specification where we interact year-specific proximity and lagged proximity with indicators for the year of instruction (pooling all years after the third). We cluster standard errors at the student level. In this setup, a student observation is included for a given year only if the student is enrolled in courses during that year. This implies the later-year estimates represent the effect of exposure specifically among the selected sample of surviving students.

These estimates indicate that frontier knowledge has the largest positive effects on degree completion and GPA when exposure happens early. Effects are strongest in the first year, then attenuate over time but remain positive. For example, a one-SD increase in proximity raises degree completion by about 5.2 pp in year 1 and 2.1 pp in year 2, compared with estimates close to zero in years 3 and 4+ (Figure 3, panel (a)). GPA effects display a similar pattern, with an estimate of about 0.11 points in year 1, 0.05 points in year 2, and smaller positive effects thereafter (panel (c)). By contrast, the effects on time-to-degree are negative in every year and become larger later in the curriculum, ranging from about -0.07 years in year 1 to roughly -0.17 years in years 4+ (panel (b)). Effects on graduate-school attendance are positive and indistinguishable from each other throughout (panel (d)). Because the later-year estimates isolate students who have successfully persisted, the attenuation in completion effects is intuitive: early-year exposure catches students when they are highly at risk of dropping out, whereas later-year exposure treats a population that has already overcome the most fragile completion margins. Overall, these estimates suggest that frontier knowledge matters most early for preventing dropout and boosting grades, while its effects on progression through the curriculum remain present throughout college.

We obtain a somewhat different pattern when looking at earnings. Frontier knowledge improves students' earnings throughout the curriculum, with somewhat larger point estimates in later years. A one-SD increase in proximity raises average earnings by about 1.6% in year 1, 2.4% in year 2, 3.6% in year 3, and 3.2% in years 4 and beyond (Figure 3, panel (e)), though the confidence intervals overlap substantially. Similarly, the same proximity increase reduces the probability of being in the bottom earnings quartile by about 1.2 pp in the first two years and about 1.9 pp in years 3 and 4+, while the probability of top-quartile earnings rises mainly in later years (panel (f)). These estimates suggest that the labor-market returns to frontier exposure are at least as large in later-stage coursework as they are earlier in college.

Taken together, these patterns suggest that frontier knowledge affects different outcomes through different margins. The largest effects on relatively fragile academic outcomes—especially degree completion and, to a lesser extent, GPA—appear when exposure occurs early in college, consistent

Figure 3: Effects of Frontier Knowledge Exposure, by Year



Note: OLS estimates; one observation is a student. The dependent variable is an indicator for students who graduate from their program (panel (a)), time-to-degree (panel (b)), GPA (panel (c)), an indicator for students ever enrolling in a graduate program within public universities in Texas (panel (d)), the natural logarithm of earnings 1 to 6 years after a student's predicted graduation year (panel (e)), and indicators for earnings in the top and bottom quartiles of a student's cohort (panel (f)). Each coefficient is an estimate of β in a stacked student-year version of equation (6) where proximity, lagged proximity, and the associated parameters are interacted with the year of instruction. Students are included in a year only if they are enrolled in courses that year. The sample is restricted to undergraduate students. Confidence intervals are calculated using standard errors clustered at the student level.

with frontier content strengthening engagement, motivation, or academic momentum at points where students are still at risk of falling off track. By contrast, the effects on higher-threshold outcomes—faster progression through the curriculum, graduate-school attendance, and earnings—are more consistent with a skill-accumulation channel: these outcomes require students not only to remain enrolled, but also to convert frontier exposure into sustained learning, stronger preparation, and downstream labor-market value. This interpretation raises the possibility that some earnings gains, especially for higher-preparation students, extend beyond measured educational attainment. We return to this point more formally in Section 5.8.

5.5 Which Courses Drive the Returns?

Another way to adjudicate between the two mechanisms is to study whether the average effects presented above are driven by one or two very high-proximity courses in a student's transcript, a result that would support a persistence and engagement channel, or by a more generalized shift in the entire curriculum, which would be more in line with a skill accumulation channel. To test for this, we calculate both the median proximity and the 90th-percentile proximity across all courses in a student's transcript and estimate the separate effects of these variables on educational outcomes within equation (6). This specification allows us to distinguish between broad exposure to frontier content and exposure driven by a small number of high-proximity courses.

Estimates from this specification on the sample of undergraduate students indicate that broad exposure matters at least as much as isolated "star" courses. A one-SD increase in median frontier knowledge proximity raises degree attainment by 2.3 pp, while a one-SD increase in 90th-percentile proximity raises it by an additional 1.8 pp (Table 4, panel (a), column 1). We observe a similar pattern for GPA (panel (a), column 3): median proximity raises GPA by 0.065 points and 90th-percentile proximity by 0.026 points. Results for Master's students point in the same direction, with positive effects of median proximity on graduation and GPA (panel (b), columns 1-3), while the corresponding estimates for PhD students are small and imprecise (panel (b), columns 4-6).

Examining effects on time-to-degree and the probability of attending graduate school points to a different margin than degree completion and GPA. For these outcomes, the strongest effects come from median proximity, with top proximity having a smaller or insignificant role. For example, a one-SD increase in median proximity, holding top proximity fixed, reduces time-to-degree by 0.19 years (panel (a), column 2) and increases the probability of graduate-school attendance by 1.2 pp (panel (a), column 4). By contrast, an increase in top proximity holding median proximity fixed has a smaller effect on time-to-degree (-0.02 years; panel (a), column 2) and essentially no effect on

Table 4: Educational Effects of Frontier Knowledge Exposure - Median vs Top Courses

Panel (a) Undergraduate students

	Graduates	Time-to degree	GPA	Attends grad school
	(1)	(2)	(3)	(4)
Median proximity (sd)	0.023***	-0.191***	0.065***	0.012***
	(0.004)	(0.017)	(0.010)	(0.004)
90th pctile proximity (sd)	0.018***	-0.018*	0.026***	0.002
	(0.003)	(0.010)	(0.006)	(0.003)
Mean dep. var.	0.873	5.669	2.885	0.116
R^2	0.562	0.595	0.451	0.287
N	190,443	155,022	189,437	190,443

Panel (b) Graduate students

	MA			PhD		
	Graduates	Time-to degree	GPA	Graduates	Time-to degree	GPA
	(1)	(2)	(3)	(4)	(5)	(6)
Median proximity (sd)	0.022**	-0.034	0.038***	0.014	-0.085	0.014
	(0.011)	(0.033)	(0.014)	(0.020)	(0.398)	(0.039)
90th pctile proximity (sd)	-0.008	-0.045**	-0.008	-0.021	0.318	0.022
	(0.007)	(0.021)	(0.010)	(0.019)	(0.269)	(0.033)
Mean dep. var.	0.714	2.294	3.355	0.257	6.302	2.810
R^2	0.509	0.696	0.516	0.820	0.891	0.787
N	26,428	21,977	26,231	4,850	1,272	4,828

Notes: OLS estimates; one observation is a student. The dependent variable is an indicator for whether the student graduated from the program (column 1 in panel (a) and columns 1 and 4 in panel (b)), time-to-degree in years (column 2 in panel (a) and columns 2 and 5 in panel (b)), GPA (column 3 in panel (a) and columns 3 and 6 in panel (b)), and an indicator for enrollment in a graduate program within Texas public universities (column 4 in panel (a)). *Median proximity* is the median frontier knowledge proximity of all courses taken by each student, measured in course-level standard deviations. *90th pctile proximity* is the proximity of the 90th percentile course in each student's transcript, also measured in course-level standard deviations. All specifications control for median and 90th-percentile lagged proximity, instructor and school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Panel (a) shows estimates for undergraduate students and panel (b) shows estimates for graduate students. Robust standard errors are reported in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

graduate-school attendance (panel (a), column 4). These results suggest that while isolated high-proximity courses may matter for more fragile outcomes such as completion and GPA, higher-threshold outcomes depend more on broad exposure to frontier content across the curriculum.

Table 5: Earnings Effects of Frontier Knowledge Exposure - Median vs Top Courses

Panel (a) Undergraduate students					
	ln(earnings)			Prob. earnings in...	
	1-6 years (1)	1-3 years (2)	4-6 years (3)	bottom qtile (4)	top qtile (5)
Median proximity (sd)	0.019**	0.026***	0.014	-0.011**	0.008*
	(0.008)	(0.010)	(0.013)	(0.005)	(0.004)
90th pctile proximity (sd)	-0.003	-0.002	-0.006	-0.002	-0.003
	(0.006)	(0.007)	(0.009)	(0.004)	(0.003)
R^2	0.292	0.296	0.350	0.211	0.311
N	190,439	158,228	86,254	190,439	190,439
Panel (b) Graduate students					
	ln(earnings)			Prob. earnings in...	
	1-6 years (1)	1-3 years (2)	4-6 years (3)	bottom qtile (4)	top qtile (5)
Median proximity (sd)	0.042***	0.056***	0.018	-0.015*	0.015
	(0.016)	(0.019)	(0.031)	(0.009)	(0.009)
90th pctile proximity (sd)	0.003	0.012	0.009	-0.001	0.004
	(0.012)	(0.013)	(0.023)	(0.006)	(0.007)
R^2	0.509	0.536	0.555	0.451	0.512
N	31,034	25,433	13,789	31,034	31,034

Notes: OLS estimates; one observation is a student. The dependent variable is the natural logarithm of earnings 1-6 years (column 1), 1-3 years (column 2), and 4-6 years after predicted graduation (column 3); an indicator for earnings in the bottom quartile (column 4) and the top quartile of the graduation cohort-specific earnings distribution (column 5). *Median proximity* is the median frontier knowledge proximity of all courses taken by each student, measured in course-level standard deviations. *90th pctile proximity* is the proximity of the 90th percentile course in each student's transcript, also measured in course-level standard deviations. All specifications control for lagged proximity (median and 90th percentile), instructor and school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Panel (a) shows estimates for undergraduate students and panel (b) shows estimates for graduate students. Standard errors in parentheses are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

The same pattern appears for earnings. Earnings effects are largely driven by median proximity, while the effects of 90th-percentile proximity are generally small and statistically indistinguishable from zero. For example, a one-SD increase in median proximity, holding top proximity fixed, raises

earnings by 1.9%, while a one-SD increase in top proximity has no detectable effect (Table 5, column 1). Similar patterns hold for the probability of being in the bottom earnings quartile (column 4) and, more modestly, for reaching the top quartile (column 5).

Taken together, these results suggest a distinction between fragile and higher-threshold outcomes. Outcomes such as degree completion and GPA can be shifted by a small number of especially frontier-oriented courses, consistent with those courses affecting engagement, motivation, or academic momentum. In contrast, outcomes that require sustained accumulation and downstream conversion of skills—faster degree progress, graduate-school attendance, and early-career earnings—are more strongly related to median proximity, that is, to broad exposure to frontier content throughout the curriculum. This pattern indicates that sustained frontier exposure—rather than exposure to one or two exceptional courses—is necessary to drive higher-threshold progression and initial labor-market placement.

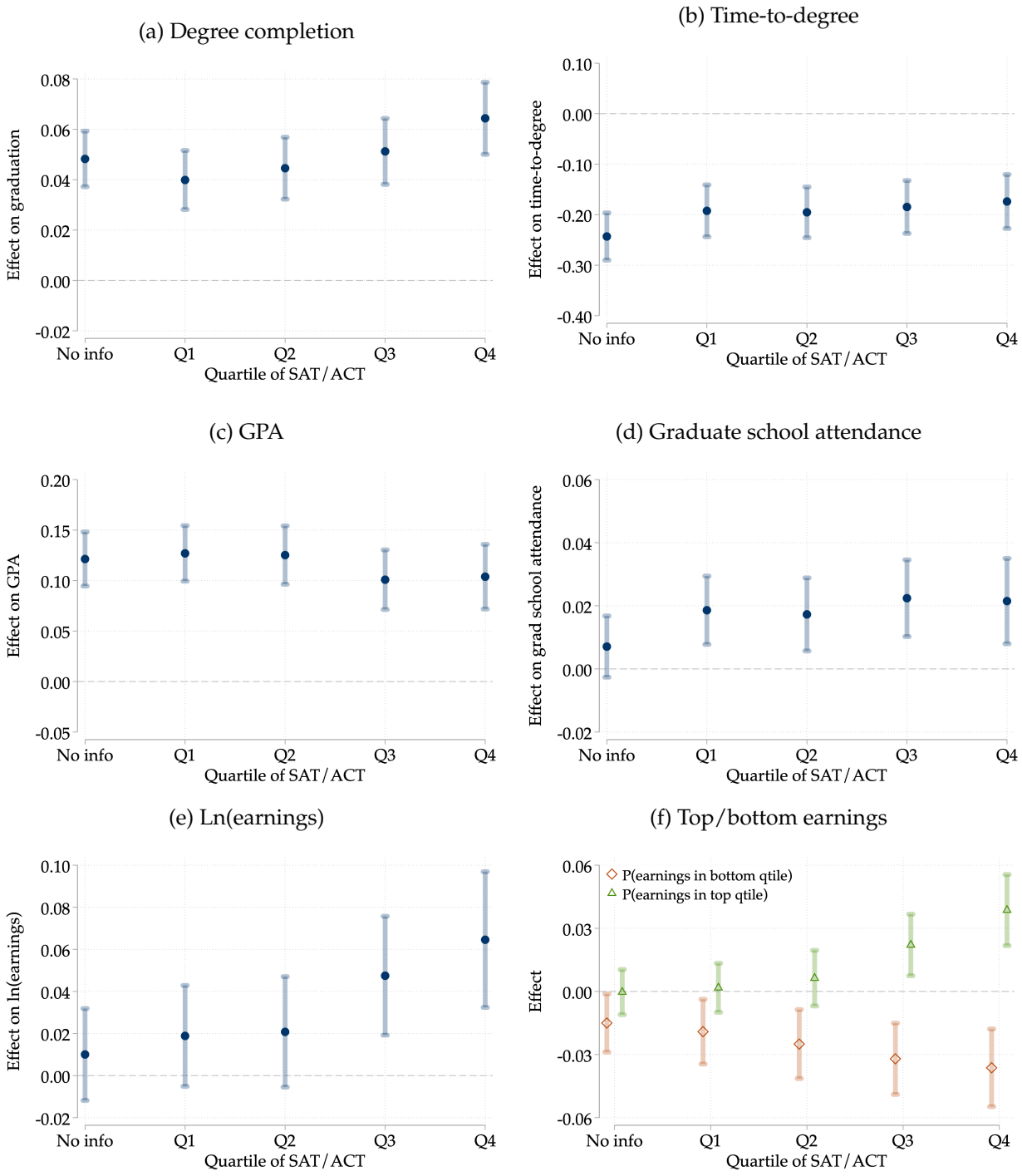
5.6 Differences by Baseline Ability

Our results so far show positive effects of frontier knowledge exposure for the average student. To understand whether these results apply to all students and to shed light on mechanisms, we now study whether these effects differ across students with different characteristics. We begin with baseline ability, by allowing the parameter β in equation (6) to vary flexibly by quartiles of the SAT/ACT score distribution. Approximately 27% of all undergraduate students in our sample do not have a reported score. This group is mainly composed of transfer students from community colleges, who are typically not required to submit a test score upon transfer. We include these students in our analysis as a separate "No Info" category.[15] This grouping lets us compare students with observed test-based ability while retaining a policy-relevant transfer population.

We first examine educational outcomes. Frontier exposure improves academic outcomes throughout the ability distribution, but the gradient differs across margins. For students in the bottom quartile of the observed ability distribution, a one-SD increase in frontier proximity raises degree completion by about 4.0 pp, reduces time-to-degree by about 0.19 years, increases GPA by about 0.13 points, and raises the likelihood of attending graduate school by about 1.9 pp (Figure 4, panels (a)-(d)). Students in the top quartile see larger improvements in degree attainment (about 6.5 pp) and graduate-school attendance, alongside similar GPA gains and somewhat smaller time-to-

[15] All four-year public institutions in our sample (and virtually all in Texas) require SAT/ACT scores for first-year admissions. Transfer students, however, need to submit scores only if they have not completed enough academic credits; the minimum credit threshold varies by school. Consequently, students without reported SAT/ACT scores almost certainly started at a two-year college, for which scores are not required. Until 2014 one of the schools in our sample, the University of Houston–Clear Lake, admitted only upper-class transfers, who were not required to provide test scores.

Figure 4: Effects of Frontier Knowledge Exposure, by Baseline Ability



Note: OLS estimates; one observation is a student. The dependent variable is an indicator for students who graduate from their program (panel (a)), time-to-degree (panel (b)), GPA (panel (c)), an indicator for students ever enrolling in a graduate program within public universities in Texas (panel (d)), the natural logarithm of earnings 1 to 6 years after a student's predicted graduation year (panel (e)), and indicators for earnings in the top and bottom quartiles of a student's cohort (panel (f)). Each coefficient is an estimate of β in a version of equation (6), allowed to vary by quartiles of the SAT/ACT scores. "No Info" refers to students without a test score. The sample is restricted to undergraduate students. Confidence intervals are calculated using robust standard errors.

degree reductions. For students in the "No Info" group, frontier exposure raises degree completion by about 4.8 pp, reduces time-to-degree by about 0.24 years, and increases GPA by about 0.12 points, but has a smaller and statistically imprecise effect on graduate-school attendance. Notably, among students with observed test scores, degree-attainment and graduate-school-attendance effects rise with ability, while time-to-degree and GPA effects are positive across the distribution and do not display the same steep gradient.

These academic patterns suggest that frontier exposure produces broad gains in performance and progression, while conversion into higher-threshold educational outcomes is more ability-complementary. GPA and time-to-degree improve across groups, indicating that lower-ability and no-score students also benefit from frontier content. The ability gradient is positive for degree completion and graduate-school attendance, suggesting that moving from improved academic performance to longer-run educational attainment requires additional preparation, information, or resources that frontier exposure alone does not fully equalize.

We next examine labor-market outcomes. The earnings effects of frontier knowledge exposure display a strong positive gradient with respect to baseline ability. While point estimates are positive across the observed ability distribution, the magnitude of the effect is much larger for students in the top ability quartile (about 6.5%) than for those in the bottom quartile (about 1.9%; Figure 4, panel (e)). For students with no reported test score, the effect on mean earnings is small and statistically indistinguishable from zero. Given that this group is largely composed of transfer students, this null is consistent with frontier exposure improving academic progression but not fully relaxing the labor-market frictions or weaker pre-college signals that may limit earnings growth.

These patterns again reveal important asymmetries. Exposure to frontier knowledge reduces the probability of low earnings throughout the observed ability distribution, with the largest reductions occurring among higher-ability students. At the upper end of the distribution, however, increases in the probability of top-quartile earnings are concentrated almost entirely among students in the top two ability quartiles. For these students, frontier exposure raises the likelihood of top-quartile earnings by about 2.3 pp in quartile 3 and 3.9 pp in quartile 4, while effects for students in the bottom half of the ability distribution and for those with no reported score are near zero and statistically insignificant (Figure 4, panel (f)).

Taken together, the ability results show a clear distinction between broad academic gains and more selective high-threshold returns. Frontier exposure improves completion, GPA, and time-to-degree across the ability distribution, and it reduces downside earnings risk among students with

observed test scores. But graduate-school attendance, mean earnings, and top-quartile earnings rise more strongly with baseline ability. Thus, the results do not imply that lower-ability students fail to benefit from frontier exposure; rather, they suggest that converting those academic gains into graduate-school and upper-tail labor-market outcomes is ability-complementary.

5.7 Differences by Family Income

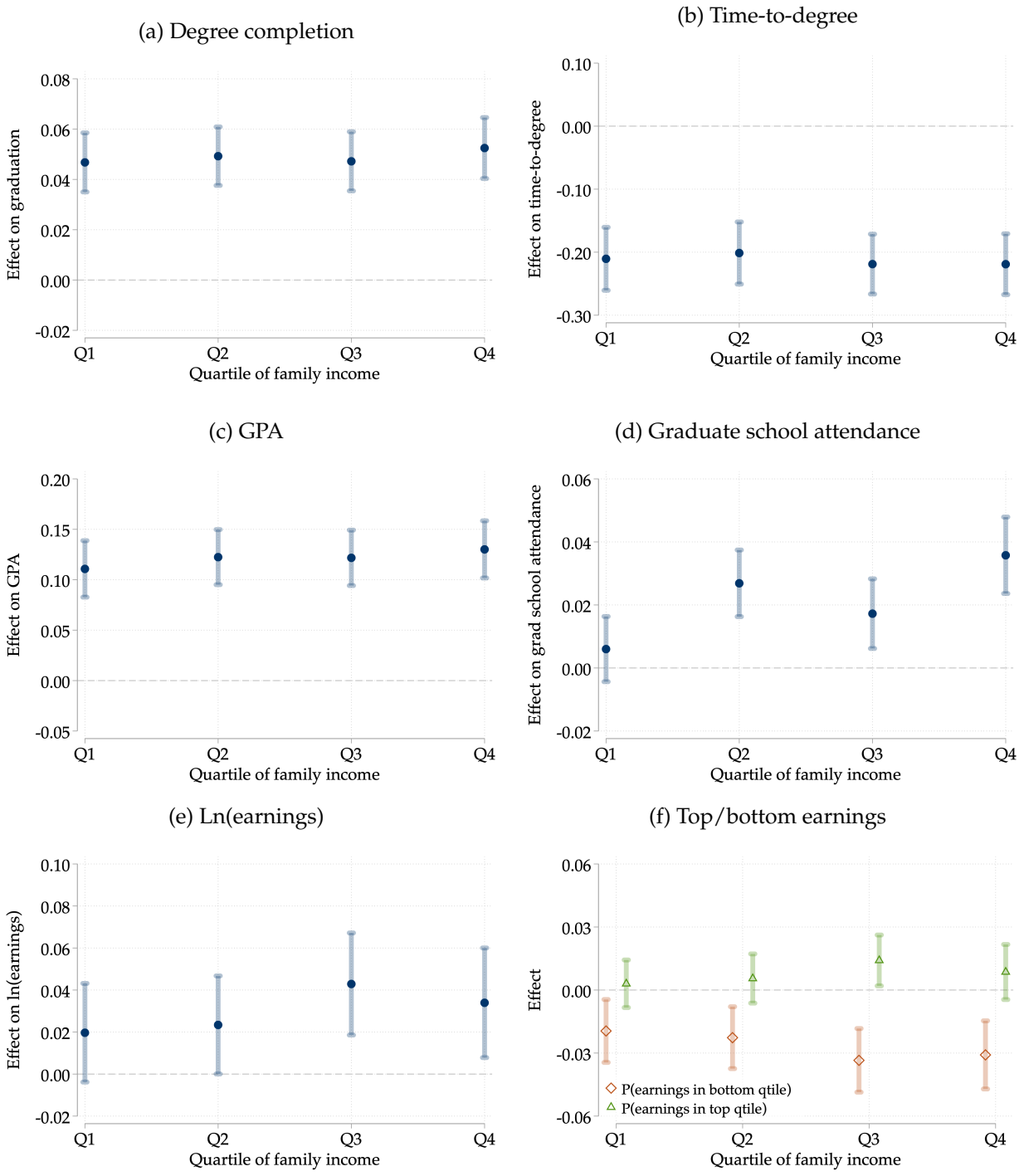
The effects of frontier knowledge exposure may also vary by socioeconomic background. For example, lower-income students may be closer to fragile completion margins, while higher-income students may have more advising, financial flexibility, and research opportunities to convert the same exposure into stronger academic and postgraduate outcomes. Testing for this type of heterogeneity thus allows us to study both distributional implications and mechanisms. To do so, we allow the parameter β in equation (6) to vary by quartiles of baseline family income (from FAFSA records). This exercise asks whether frontier knowledge exposure is equally productive across students with different pre-college resources.

As with ability, our results indicate that frontier knowledge helps students across the family income distribution, but along different margins. Undergraduate students in all income quartiles experience sizable gains in degree completion (Figure 5, panel (a)) and faster time-to-degree (panel (b)), with no clear income gradient in either outcome. GPA effects are also positive throughout the distribution and rise modestly with income (panel (c)). Graduate-school attendance instead displays a sharper gradient: the effect is small and statistically indistinguishable from zero in the bottom income quartile, but positive above it and largest in the top quartile (panel (d)).

We next examine labor-market returns by undergraduate students' socioeconomic backgrounds. The earnings point estimates are positive across the family income distribution, but they are larger and more precisely estimated in the upper half of the distribution (Figure 5, panel (e)). This pattern suggests that students from higher-income families are better positioned to translate frontier exposure into earnings gains in the labor market. The top- and bottom-quartile earnings outcomes reinforce this interpretation. Exposure to frontier knowledge reduces the probability of having earnings in the bottom quartile across all family income groups, with somewhat larger reductions in the upper half of the income distribution. In contrast, increases in the probability of having earnings in the top quartile are concentrated among students from the top half of the family-income distribution, though these estimates are more modest and less precise (Figure 5, panel (f)).

Taken together, these results suggest that frontier exposure improves completion and academic progression broadly across the income distribution. However, the conversion of frontier exposure

Figure 5: Effects of Frontier Knowledge Exposure, by Family Income



Note: OLS estimates; one observation is a student. The dependent variable is an indicator for students who graduate from their program (panel (a)), time-to-degree (panel (b)), GPA (panel (c)), an indicator for students ever enrolling in a graduate program within public universities in Texas (panel (d)), the natural logarithm of earnings 1 to 6 years after a student's predicted graduation year (panel (e)), and indicators for earnings in the top and bottom quartiles of a student's cohort (panel (f)). Each coefficient is an estimate of β in a version of equation (6), allowed to vary by quartiles of the family income distribution. The sample is restricted to undergraduate students. Confidence intervals are calculated using robust standard errors.

into higher-threshold outcomes is resource-dependent: GPA gains are somewhat larger at higher income levels, graduate-school gains are concentrated outside the bottom quartile, and average earnings gains are larger in the upper half of the income distribution. From an inequality perspective, frontier exposure appears to improve fragile academic margins and reduce downside earnings risk broadly, while graduate-school progression and higher-return labor-market outcomes remain more closely tied to family resources.

5.8 Decomposing the Earnings Effects: Mediation Analysis

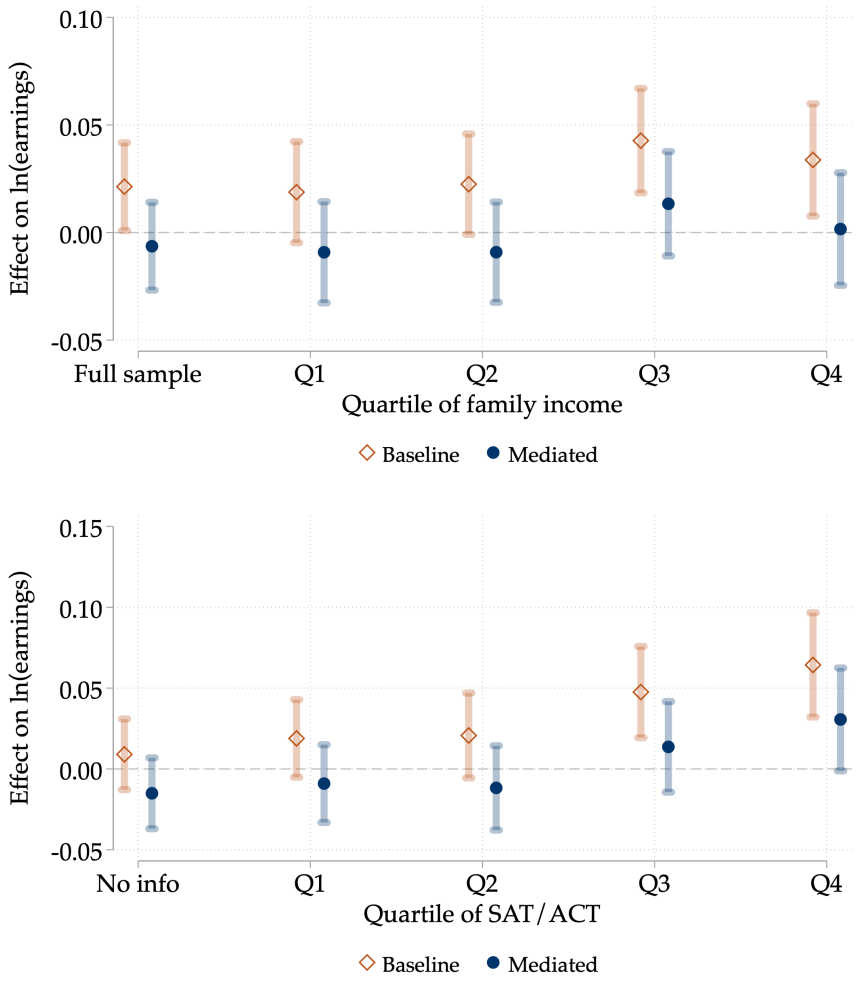
Our results so far indicate that exposure to frontier knowledge in college improves both educational outcomes and earnings. These gains, however, differ systematically across student groups. Frontier exposure improves academic and progression margins such as completion, GPA, and time-to-degree broadly across income and ability groups, including among students without reported test scores. By contrast, graduate-school attendance, mean earnings, and upper-tail earnings tend to rise more with baseline ability and family resources. This pattern suggests that earnings gains may combine two channels: improvements in observed educational progression and performance, and additional labor-market returns to frontier-acquired skills that are more complementary with students' pre-college resources.

To discipline this interpretation and, more generally, to better understand how much of the earnings gains can be accounted for by improvements in educational attainment and performance, we conduct a descriptive mediation analysis. We contrast our baseline earnings estimates with estimates from specifications that additionally control for degree completion, graduate school enrollment, cumulative GPA, and time-to-degree. When examining heterogeneous effects by family income and ability, we interact these educational controls with the relevant group indicators. Because these controls are themselves outcomes affected by frontier exposure, we interpret the exercise descriptively, as an accounting decomposition rather than as a causal mediation estimate.

The results from this exercise are shown in Figure 6. In the full sample, the coefficient on frontier knowledge proximity falls from 2.1% in the baseline specification to -0.6% after controlling for educational outcomes, and the mediated estimate is statistically indistinguishable from zero. This pattern indicates that, on average, the earnings return to frontier exposure is largely accounted for by differences in observed educational attainment, progression, and performance.

The attenuation is especially pronounced for lower-resource and lower-preparation groups. Among students in the bottom two family-income quartiles, the mediated estimates are slightly negative and statistically indistinguishable from zero. The same is true for students with no re-

Figure 6: Earnings Effects of Frontier Knowledge Exposure: Mediation Analysis



Notes: The figure shows *baseline* estimates of frontier knowledge proximity on the natural logarithm of earnings 1-6 years after predicted graduation, along with the same *mediated* estimates—obtained controlling for intermediate educational outcomes (graduation, time-to-degree, GPA, and an indicator for graduate school attendance). The top panel shows estimates in the full sample and by quartile of family income, obtained allowing β in equation (6) to vary across quartiles. The bottom panel shows estimates by quartile of the ability distribution, allowing β in equation (6) to vary across quartiles. Mediated estimates by family income and ability are obtained by also interacting intermediate educational outcomes with indicators for the relevant groups. Standard errors are clustered at the student level.

ported test score and for students in the bottom half of the observed ability distribution. These results are consistent with the earnings gains for these groups operating mainly through observable educational margins: completing the degree, progressing faster, earning higher grades, or enrolling in graduate school.

For more advantaged students, the evidence is more suggestive but points to a different margin. By family income, the mediated estimate remains positive only in the third quartile, at about 1.3%, while the top-quartile mediated estimate is close to zero. By ability, the mediated estimates remain

positive for the top two quartiles and are largest in the top quartile, at about 3.1%, though these estimates are not statistically distinguishable from zero. Thus, the decomposition is consistent with frontier exposure improving educational progression broadly, but the part of earnings gains that is not absorbed by observed educational outcomes is concentrated among students with stronger baseline preparation, with a weaker and less monotonic pattern by family income.

6 Robustness

In this section, we probe the robustness of our results by assessing three issues: (i) whether our identifying variation is plausibly conditionally exogenous, (ii) whether selective add/drop behavior around the start of the semester contaminates our estimates, and (iii) whether correlated changes in course design or instructor workload can account for our results.

6.1 Conditional Independence: Placebo Test with Future Exposure

Our identification strategy assumes that, conditional on lagged proximity and fixed effects, the realized frontier proximity of the courses a student takes is unrelated to unobserved determinants of the student's outcomes. A natural concern is that students with stronger unobserved preparation, motivation, or information both perform better and end up taking courses with higher frontier proximity. If this type of selection remains after our controls, it should be visible even before the relevant exposure is realized: students who will later take higher-proximity courses should already rank higher in the courses they are currently taking.

We test this implication using a placebo regression at the student-course level. For each student i and focal course k, we regress the student's grade rank in course k on the average frontier knowledge proximity of all courses that student i takes *after* course k, measured in course-level standard deviations. We further control for the lagged proximity of these courses, mirroring our main specification, as well as student controls, instructor fixed effects, and school-major-cohort fixed effects. The intuition behind this test relies on a combination of temporal ordering and our conditional independence assumption. By definition, future course content cannot causally affect past performance. Therefore, if our identifying assumptions hold—meaning there is no residual selection on unobservables—the estimated coefficient on future proximity must be zero. A nonzero coefficient would instead indicate that future frontier exposure is proxying for unobserved student characteristics or course-taking trajectories that predict current academic performance.

The results are reported in Appendix Table A7. The estimated coefficient on future proximity is statistically insignificant and economically negligible both for undergraduates (-0.0004, with a

standard error of 0.0012) and for graduate students (-0.0026, with a standard error of 0.0023). The absence of any relationship between future frontier exposure and current grades supports the view that, conditional on our controls, frontier proximity is uncorrelated with latent student ability, motivation, or course-path selection.

6.2 Selective Add/Drop and Late Enrollment

A distinct threat to identification arises from the add/drop window that opens once a semester begins and the updated syllabus becomes publicly visible. Students who drop an initial enrollment and add a different course during this window could be selecting on the realized proximity update $\Delta_{k,t}$, precisely the variation we rely upon for identification.

We address the dropping concern at the estimation stage by assigning each course to a student based on initial enrollment regardless of subsequent completion, yielding intent-to-treat (ITT) estimates. Late adds are a subtler concern because we cannot directly observe post-registration enrollment dates. However, since adding a course after the registration deadline generally requires dropping another, students who drop any course in a given year constitute the pool most likely to include late adders. We therefore re-estimate our main specifications on the sample of students who did not drop any course in a given academic year. As shown in Appendix Tables A12 and A13, the effects on educational and earnings outcomes are virtually unchanged relative to the baseline, indicating that selective add/drop behavior does not materially affect our estimates.

6.3 Time-Varying Instructor Effects

A potential threat to our strategy is that changes in a course's frontier knowledge proximity may be correlated with time-varying instructor effects that directly affect student outcomes. For example, consider an instructor who has been teaching the same course, with the same material, for many years and suddenly decides to update the course. Such an update could coincide with an improvement in teaching effectiveness. If this occurrence is prevalent, it could explain part of the estimates we find. More generally, we want to account for the possibility that an increase in instructor effort could simultaneously lead to greater incorporation of frontier material and improvements in teaching effectiveness.

While we are unable to directly observe teaching ability beyond instructor fixed effects, we probe this threat in two ways: by letting instructor effects vary semi-parametrically with experience, and by excluding large course updates, i.e., dramatic changes in course content that could coincide with large changes in instructor effort.

To implement the first test, we augment equation (6) by flexibly controlling for instructors' teaching experience, a proxy for time-varying teaching effectiveness. We define teaching experience as the number of years elapsed between the year of course instruction and the year in which the instructor first appears teaching in our data. We then augment equation (6) to include fixed effects for the instructor's experience. As shown in Appendix Tables A16 and A17, our main estimates are largely unchanged.[16]

To conduct the second test, we calculate p_i and $p_{i,-1}$ in equation (6) by excluding courses that experience a major content overhaul, defined as a proximity update larger than one SD (or 27% of the course's content). Our estimates are largely unchanged (Appendix Tables A10 and A11).

Notably, estimates that exclude both instructor fixed effects and experience controls are similar in magnitude to our baseline estimates (Appendix Tables A8 and A9). Although this is a less preferred specification, it further indicates that instructor-specific effects do not appear to be systematically related to changes in course proximity.

6.4 Other Dimensions of Course Content and Assessment

A distinct concern is that frontier knowledge updates may co-occur with changes in other aspects of course design—in particular, assessment structure—that directly affect student learning and outcomes, independently of the knowledge content itself. For instance, a course that incorporates more cutting-edge material might simultaneously shift its evaluation from traditional exams toward research reports or presentations. Such a shift could develop different skill sets and human capital, and therefore affect earnings or degree completion for reasons unrelated to proximity. In this case, our estimates would capture a bundle of course design changes rather than the effect of frontier content alone.

To assess this possibility, we augment equation (6) with controls for syllabus grading structure—specifically, the shares of total course assessment allocated to exams, homework, and reports, captured by the shares of corresponding terms in the assignments/grading portion of a syllabus. As shown in Appendix Tables A14 and A15, the estimates of the effect of frontier knowledge proximity on educational and earnings outcomes are only slightly attenuated and still economically and statistically significant, indicating that correlated changes in assessment practices do not account for our main findings.

[16] In Appendix Tables A16 and A18, we omit estimates on PhD students' time-to-degree due to a lack of sufficient sample size and variation to identify these effects.

6.5 Instructor Teaching Load

A related but conceptually distinct concern is that proximity updates may be correlated with changes in instructors' workload that affect the quality of instruction delivered. For example, an instructor teaching fewer sections in a given semester has more time to invest in each course, potentially improving teaching quality simultaneously with updating course content. If this pattern is widespread, part of our estimates could reflect improved instruction rather than greater frontier knowledge exposure per se.

To probe this, we augment equation (6) with controls for instructors' teaching load, measured by the number of sections taught in the same year. As shown in Appendix Tables A18 and A19, results are largely unchanged, suggesting that correlated variation in instructors' workload does not drive our findings.

7 Discussion and Conclusion

This paper has shown that proximity to frontier knowledge is an important dimension of college course content. Greater exposure to courses closer to the knowledge frontier improves educational outcomes—degree completion, academic performance, and graduate-school attendance—and raises earnings. However, the returns to frontier exposure are not uniform across outcome margins or student groups. Frontier exposure improves degree completion and academic progression broadly, while its effects on graduate-school attendance and earnings are stronger among students with greater pre-college resources and preparation.

The evidence points to two complementary channels for these effects. Exposure in the first year has the largest effects on completion and GPA, consistent with frontier content strengthening engagement, motivation, and academic momentum. Broader exposure across the curriculum is strongly related to faster progression, graduate-school attendance, and short-run earnings. For the average student, these early earnings gains largely reflect improved degree completion and initial labor-market placement. For higher-ability students, however, some of the earnings premium remains after conditioning on attainment, consistent with additional returns of cumulative skill accumulation.

These findings suggest that curriculum content, and specifically proximity to frontier knowledge, is a policy-relevant input into human capital production. Because substantial variation in frontier exposure arises across both courses and instructors, universities may affect student outcomes not only through program offerings, institutional resources, or admissions policies, but also

through the content of the courses students encounter, especially in introductory, gateway, and required classes. In a companion paper (Biasi and Ma, 2022), we use a broader national sample of syllabi to show that research-active instructors—those with more publications, citations, and grants, and whose research closely matches the course topic—teach courses with higher frontier proximity. Combined with the evidence in this paper, that pattern suggests a complementarity between research and teaching: faculty research activity is associated with more frontier-oriented instruction, and such instruction improves student outcomes.

When interpreting these policy implications, it is important to note the specific nature of our estimand. Because our empirical design relies on unobserved, term-to-term syllabus updates, our estimates capture the effects of localized, instructor-led modernizations rather than holistic, program-wide curriculum overhauls. While we show that these surprise updates deliver substantial benefits to students, a coordinated institutional effort to intentionally modernize a department's entire required curriculum could theoretically yield different—and potentially larger—cumulative effects.

An important boundary on our conclusions stems from the coverage of our administrative data. While our estimates are representative of students who enter Texas public graduate programs or the Texas UI-covered labor market, we cannot observe graduate enrollment outside the Texas public system or earnings for students who leave the state or enter employment not covered by UI records. Determining whether the returns to frontier exposure are similar, larger, or smaller for students who migrate out of state or enroll in other graduate institutions remains an important avenue for future research. The data and methodology developed here also open avenues for studies of other dimensions of course content—problem-set structure, interdisciplinarity, the integration of empirical methods—and outcomes beyond those studied here, including innovation and patent production. Whether differences in frontier exposure translate into differences in the creation of new ideas is a consequential next question.

References

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, Volume 4, pp. 1043–1171. Elsevier.

Akcigit, U., J. G. Pearce, and M. Prato (2025). Tapping into talent: Coupling education and innovation policies for economic growth. *Review of Economic Studies 92*(2), 696–736.

Altonji, J. G., E. Blom, and C. Meghir (2012, sep). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics 4*(1), 185–223.

American Association of University Professors (1940). 1940 Statement of Principles on Academic Freedom and Tenure. *AAUP Bulletin 64*(2), 108–112.

Andrews, M. J. (2023). How do institutions of higher education affect local invention? Evidence from the establishment of US colleges. *American Economic Journal: Economic Policy 15*(2), 1–41.

Angrist, J., P. Azoulay, G. Ellison, R. Hill, and S. F. Lu (2017). Economic research evolves: Fields and styles. *American Economic Review 107*(5), 293–97.

Arold, B. W. (2024). Evolution vs. Creationism in the classroom: The lasting effects of science education. *The Quarterly Journal of Economics 139*(4), 2331–2375.

Arteaga, C. (2018). The effect of human capital on earnings: Evidence from a reform at Colombia's top university. *Journal of Public Economics 157*, 212–225.

Aryal, G., M. Bhuller, and F. Lange (2022). Signaling and employer learning with instruments. *American Economic Review 112*(5), 1669–1702.

Baumol, W. J. (2005). Education for innovation: Entrepreneurial breakthroughs versus corporate incremental improvements. *Innovation Policy and the Economy 5*, 33–56.

Bianchi, N. and M. Giorcelli (2020). Scientific education and innovation: from technical diplomas to university STEM degrees. *Journal of the European Economic Association 18*(5), 2608–2646.

Biasi, B. and S. Ma (2022). The Education-Innovation Gap. NBER Working Paper 29853, National Bureau of Economic Research.

Biasi, B. and P. Moser (2021). Effects of copyrights on science: Evidence from the WWII book republication program. *American Economic Journal: Microeconomics 13*(4), 218–60.

Cantoni, D., Y. Chen, D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2017). Curriculum and ideology. *Journal of Political Economy 125*(2), 338–392.

Cunha, J. M. and T. Miller (2014). Measuring value-added in higher education: Possibilities and limitations in the use of administrative data. *Economics of Education Review 42*, 64–77.

Dale, S. B. and A. 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.

Dale, S. B. and A. B. Krueger (2014). Estimating the effects of college characteristics over the career using administrative earnings data. *Journal of Human Resources 49*(2), 323–358.

Dee, T. S. and E. K. Penner (2017). The causal effects of cultural relevance: Evidence from an ethnic studies curriculum. *American Educational Research Journal 54*(1), 127–166.

Deming, D. J. and K. L. Noray (2020). Earnings dynamics, changing job skills, and STEM careers. *Quarterly Journal of Economics 135*(4), 1965–2005.

Galasso, A. and M. Schankerman (2015). Patents and cumulative innovation: Causal evidence from the courts. *The Quarterly Journal of Economics 130*(1), 317–369.

Goldin, C. D. and L. F. Katz (2010). *The Race Between Education and Technology*. Harvard University Press.

Hemelt, S. W., B. Hershbein, S. M. Martin, and K. M. Stange (2023). College majors and skills: Evidence from the universe of online job ads. *Labour Economics 85*, 102429.

Hoxby, C. M. (1998). The return to attending a more selective college: 1960 to the present. *Unpublished manuscript, Department of Economics, Harvard University, Cambridge, MA*.

Hoxby, C. M. and G. Bulman (2015). Computing the value-added of American postsecondary institutions. Mimeo, Stanford University.

Huettner, F., M. Sunder, et al. (2012). REGO: Stata module for decomposing goodness of fit according to Owen and Shapley values. In *United Kingdom Stata Users' Group Meetings 2012*. Stata Users Group.

Iaria, A., C. Schwarz, and F. Waldinger (2018). Frontier knowledge and scientific production: evidence from the collapse of international science. *The Quarterly Journal of Economics 133*(2), 927–991.

Israeli, O. (2007). A Shapley-based decomposition of the R-square of a linear regression. *The Journal of Economic Inequality 5*(2), 199–212.

Joensen, J. S. and H. S. Nielsen (2009). Is there a causal effect of high school math on labor market outcomes? *Journal of Human Resources 44*(1), 171–198.

Jones, B. F. (2009). The burden of knowledge and the death of the Renaissance man: Is innovation getting harder? *Review of Economic Studies 76*(1), 283–317.

Kelly, B., D. Papanikolaou, A. Seru, and M. Taddy (2021). Measuring technological innovation over the long run. *American Economic Review: Insights 3*(3), 303–20.

Li, X., S. Linde, and H. Shimao (2021). Major Complexity Index and college skill production. *Available at SSRN 3791651*.

Ma, X. and E. Hovy (2016). End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, pp. 1064–1074. Association for Computational Linguistics.

Moser, P. and A. Voena (2012). Compulsory licensing: Evidence from the Trading with the Enemy Act. *American Economic Review 102*(1), 396–427.

Mountjoy, J. and B. R. Hickman (2021). The returns to college(s): Relative value-added and match effects in higher education. NBER Working Paper 29276, National Bureau of Economic Research.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy 94*(5), 1002–1037.

Toivanen, O. and L. Väänänen (2016). Education and invention. *Review of Economics and Statistics 98*(2), 382–396.

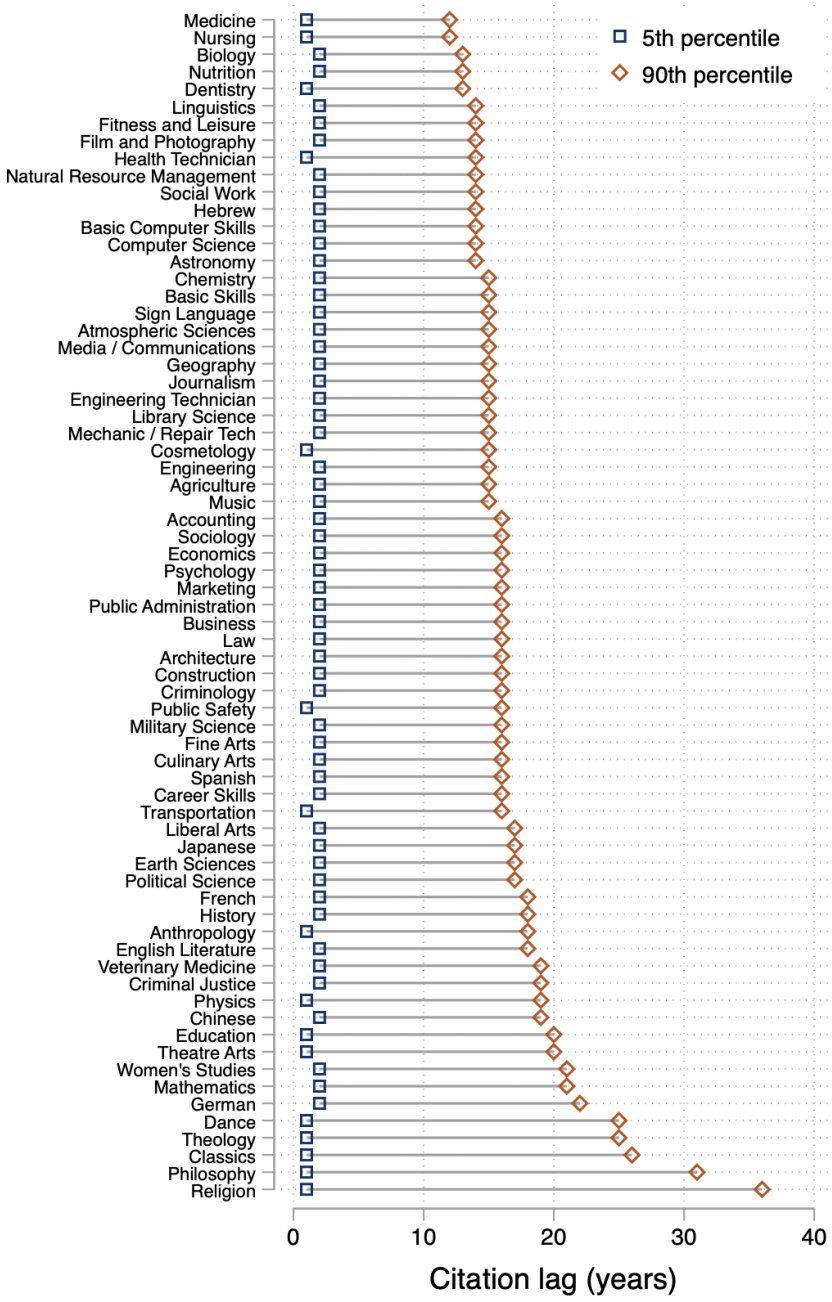
Williams, H. L. (2013). Intellectual property rights and innovation: Evidence from the human genome. *Journal of Political Economy 121*(1), 1–27.

Appendix

For online publication only

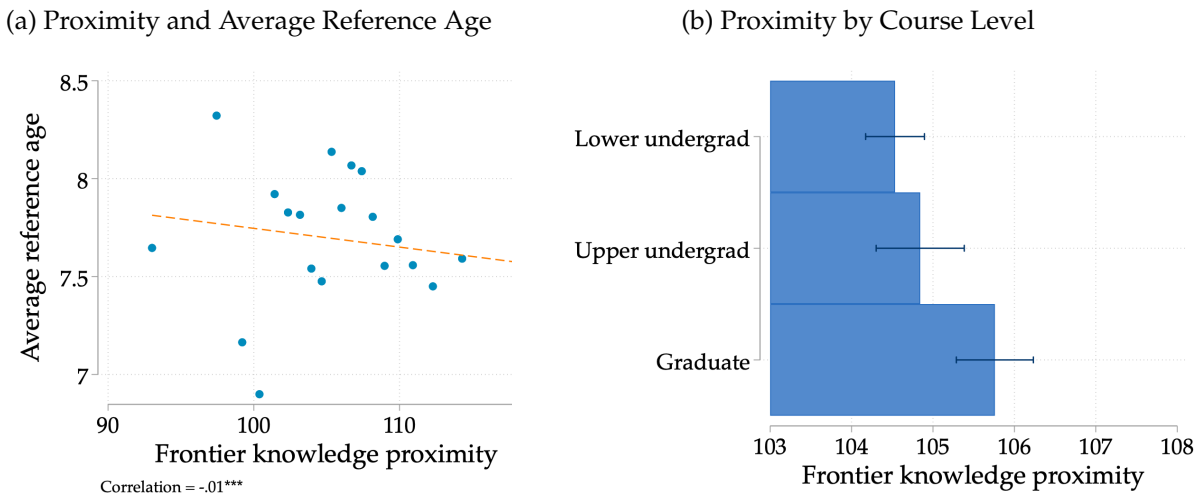
Appendix A Additional Tables and Figures

Figure A1: Citation Lags by Field: 5th and 90th Percentiles



Notes: 5th and 90th percentile of the citation lag distribution in each field, used to calculate the frontier knowledge proximity for the syllabi in each field.

Figure A2: Validating Frontier Knowledge Proximity



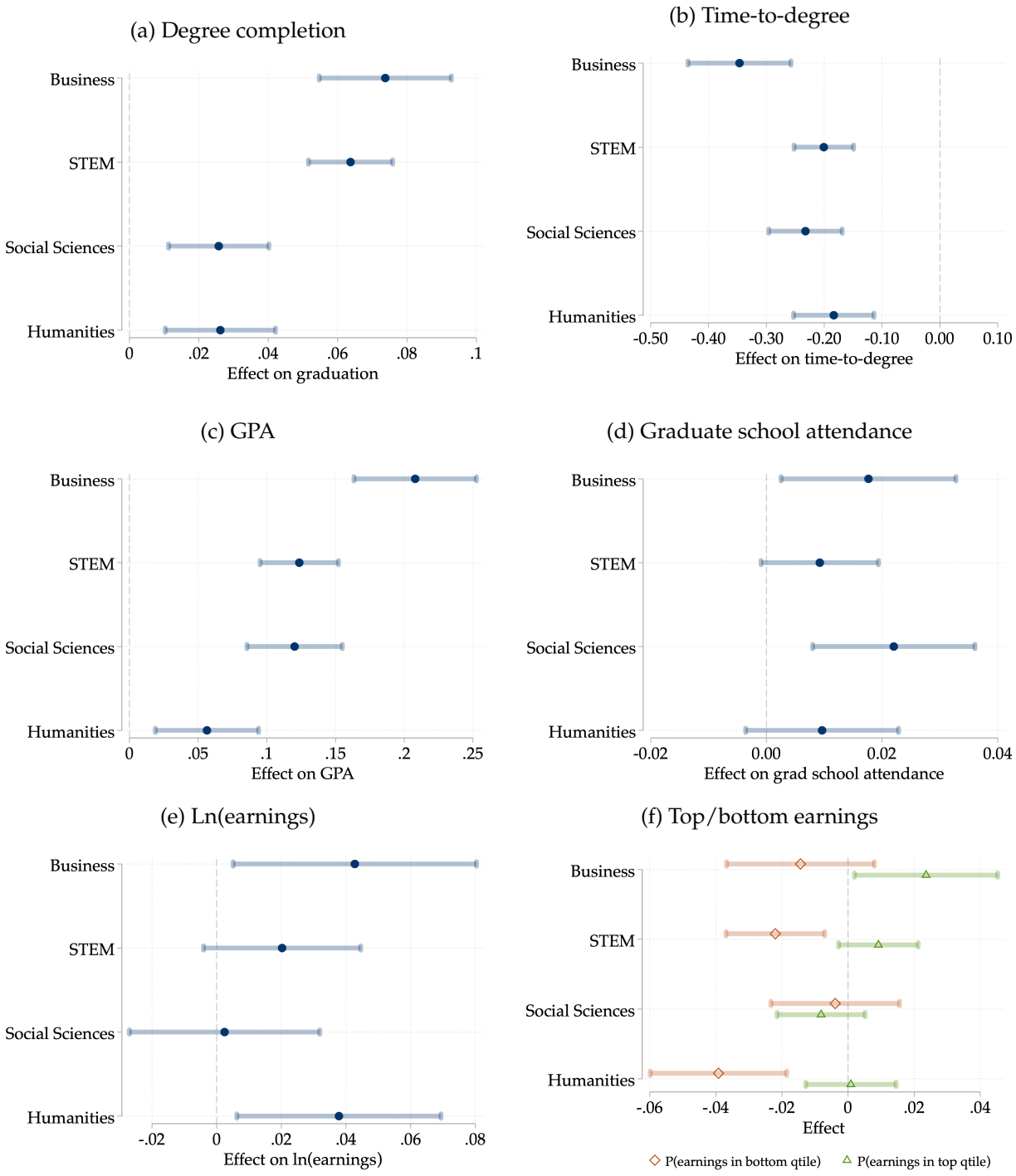
Note: Panel (a) shows a binned scatterplot of the proximity to the knowledge frontier and the average age of a syllabus's references (required or recommended readings), where reference age is calculated as the difference between the year of the syllabus and the year of publication of each reference. Panel (b) shows the mean and 95-percent confidence intervals of the proximity by course level, controlling for field-by-year effects.

Table A1: Decomposing the Variation in the Frontier Knowledge Proximity: Schools, Years, Fields, Courses, and Instructors

	Partial R^2	
	(1)	(2)
Year	0.087	0.079
Field	0.132	0.046
School	0.007	0.003
Course	.	0.258
Instructor	.	0.210
All	0.226	0.597

Note: This table shows a Shapley-Owen decomposition of the adjusted R^2 of a regression of proximity on fixed effects for schools, years, fields, instructors, and courses into the contribution of each set of fixed effects. *All* reports the adjusted R^2 of a regression with all sets of fixed effects included. We use adjusted R^2 in lieu of R^2 to account for the large number of fixed effects.

Figure A3: Effects of Frontier Knowledge Exposure, by Field



Note: OLS estimates; one observation is a student. The dependent variable is an indicator for students who graduate from their program (panel (a)), time-to-degree (panel (b)), GPA (panel (c)), an indicator for students ever enrolling in a graduate program within public universities in Texas (panel (d)), the natural logarithm of earnings 1 to 6 years after a student's predicted graduation year (panel (e)), and indicators for earnings in the top and bottom quartiles of a student's cohort (panel (f)). Each coefficient is an estimate of β in a version of equation (6), allowed to vary by macro-field. The sample is restricted to undergraduate students. Confidence intervals are calculated using robust standard errors.

Table A2: Timeline of Course Enrollment, Add/Drop Decisions, and Syllabi Visibility - Texas Sample

School	Registration Period	Date Checked	Syllabi Visible?	Add/Drop Deadline
UT Austin	Dec 16 - Dec 20	Dec 16	No	Jan 30
Texas A&M	Nov 18	Nov 18	No	Jan 10
UT Dallas	Oct 21	Nov 20	No	Jan 26
West Texas A&M	Nov 6- Nov 11	Nov 20	Very few visible	Jan 28
Sam Houston State	Oct 31-Nov 14	Nov 19	No	Jan 18
Stephen F Austin	Oct 30 - Nov 4	Nov 19	No	Jan 24
U Houston Clear Lake	Nov 20	Nov 20	No	Feb 1

Note: The table summarizes the timeline of course enrollment, add/drop decisions, and syllabi visibility for the Winter/Spring semester of 2025 at the seven universities in our sample.

Table A3: Predicting Changes in Course Proximity Using Observables

	Proximity ($p_{k,t}$)		Change in proximity ($\Delta_{k,t}$)	
Variable	(1)	(2)	(3)	(4)
Family income in 2nd quartile	-0.012	-0.013	-0.022	-0.034*
	(0.020)	(0.023)	(0.019)	(0.020)
Family income in 3rd quartile	0.001	0.018	-0.037*	-0.055***
	(0.021)	(0.023)	(0.019)	(0.020)
Family income in 4th quartile	0.020	0.031	-0.018	-0.027
	(0.021)	(0.023)	(0.019)	(0.020)
Missing family income	0.062***	0.026	-0.007	-0.008
	(0.024)	(0.027)	(0.023)	(0.025)
SAT/ACT in 1st quartile	-0.090***	-0.040*	0.004	-0.005
	(0.018)	(0.021)	(0.017)	(0.018)
SAT/ACT in 2nd quartile	-0.141***	-0.102***	0.028	0.041**
	(0.019)	(0.022)	(0.018)	(0.019)
SAT/ACT in 3rd quartile	-0.104***	-0.079***	0.028	0.034
	(0.020)	(0.023)	(0.021)	(0.021)
SAT/ACT in 4th quartile	-0.104***	-0.097***	0.010	0.015
	(0.021)	(0.025)	(0.021)	(0.022)
Share female	-0.007	-0.025*	0.001	-0.002
	(0.013)	(0.015)	(0.013)	(0.013)
Courses/Syllabi				
Grade points		0.127***		-0.004
		(0.007)		(0.006)
Requires exam		-0.160***		
		(0.037)		
ΔRequires exam				-0.080*
				(0.042)
Share homework		0.052		
		(0.040)		
ΔShare homework				0.093
				(0.062)
Share report		0.057**		
		(0.023)		
ΔShare report				-0.002
				(0.040)
Instructors				
Nr instructors = 2		-0.027***		0.020**
		(0.008)		(0.008)
Nr instructors = 3+		-0.061***		0.020**
		(0.009)		(0.009)
Experience 2-3 years		0.017		-0.013
		(0.017)		(0.018)
Experience 4-5 years		-0.005		-0.022
		(0.018)		(0.019)
Experience 6+ years		-0.026		-0.027
		(0.018)		(0.019)
New instructor		-0.015**		-0.011
		(0.008)		(0.008)
N (course * year)	102,785	73,972	74,657	62,648
F-stat p-value, student variables	0.00	0.00	0.67	0.10
F-stat p-value, syllabi variables		0.00		0.13
F-stat p-value, instructor variables		0.00		0.09

Note: OLS estimates; one observation is a course-year. The dependent variable is frontier knowledge proximity (columns 1 and 2) and the change in proximity from the previous year (columns 3 and 4). The independent variables are average characteristics of all students taking the course, of the syllabus, and of the course's instructors. *Grade points* is the average grade of students taking the course. *Exam* equals one if the course requires an exam; *Share homework* and *Share report* are the shares of terms in the assignment portion of a syllabus that refer to these activities; and variables starting with Δ denote changes to the variable from the previous year. Instructor experience is calculated as the time elapsed since the instructor first appears teaching in the data. All specifications control for field (as indicated by the course prefix) by-school-by-year fixed effects. Standard errors in parentheses are clustered at the course level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A4: Frontier Knowledge Exposure and Appearance in the Earnings Sample

	Undergraduate		Graduate	
	Missing earnings (1)	3+ cons. earnings qts (2)	Missing earnings (3)	3+ cons. earnings qts (4)
Proximity (sd)	-0.005	0.003	-0.005	0.001
	(0.004)	(0.003)	(0.005)	(0.003)
Mean dep. var.	0.466	0.075	0.400	0.097
R^2	0.170	0.163	0.325	0.254
N	409,136	409,136	87,274	87,274

Notes: OLS estimates on the full sample of students, regardless of the presence of earnings information; one observation is a student. Columns 1-2 report results for undergraduate students and columns 3-4 for graduate students. The dependent variable is an indicator for having no earnings information (columns 1 and 3) or for having non-missing earnings information for three or more consecutive quarters (columns 2 and 4); both are measured 1-6 years after predicted graduation. *Proximity* is the average frontier knowledge proximity of the courses taken by each student, measured in course-level standard deviations. All specifications control for average lagged proximity, instructor fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A5: Educational Effects of Frontier Knowledge Exposure - Full Sample of Students

Panel (a) Undergraduate students					
	Graduates	Time-to degree	GPA	Attends grad school	Future grade rank
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.044***	-0.105***	0.089***	0.010***	0.005***
	(0.003)	(0.016)	(0.008)	(0.003)	(0.001)
Mean dep. var.	0.825	5.603	2.916	0.102	0.524
R^2	0.579	0.556	0.405	0.248	0.085
N	409,670	305,305	407,149	409,670	3,674,033
N clusters (students)	.	.	.	.	162,091

Panel (b) Graduate students						
	MA			PhD		
	Graduates	Time-to degree	GPA	Graduates	Time-to degree	GPA
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.007	-0.050***	0.028***	-0.003	-0.118*	0.028*
	(0.006)	(0.018)	(0.008)	(0.008)	(0.061)	(0.015)
Mean dep. var.	0.698	2.159	3.347	0.287	5.650	2.896
R^2	0.424	0.612	0.427	0.602	0.727	0.653
N	73,636	54,652	73,156	21,159	7,296	21,369

Notes: OLS estimates on the full sample of students, regardless of the presence of earnings information. Panel (a) reports results for undergraduate students and panel (b) for graduate students. The dependent variable is degree completion (column 1 in panel (a) and columns 1 and 4 in panel (b)), time-to-degree in years (column 2 in panel (a) and columns 2 and 5 in panel (b)), GPA (column 3 in panel (a) and columns 3 and 6 in panel (b)), graduate school attendance within Texas public universities (column 4 in panel (a)), and the student's rank in future course grades (column 5 in panel (a)). In all columns except column 5 of panel (a), one observation is a student; in column 5 of panel (a), one observation is a student-course pair. *Proximity* is the average frontier knowledge proximity of the courses taken by each student, measured in course-level standard deviations; in column 5 of panel (a), it is calculated using only courses taken before the focal course. All specifications control for average lagged proximity, instructor fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses in all columns except column 5 of panel (a), where standard errors are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A6: Educational Effects of Frontier Knowledge Exposure - Sample of Undergraduate Students Who Reach Year 3

	Graduates	Time-to degree	GPA	Attends grad school	Future grade rank
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.042***	-0.213***	0.075***	0.010*	0.004***
	(0.006)	(0.024)	(0.013)	(0.005)	(0.001)
Mean dep. var.	0.904	5.659	2.940	0.120	0.524
R^2	0.463	0.606	0.394	0.292	0.085
N	172,035	150,570	171,470	172,035	3,673,834

Notes: OLS estimates on the sample of undergraduate students who reach year 3. The dependent variable is degree completion (column 1), time-to-degree in years (column 2), GPA (column 3), graduate school attendance within Texas public universities (column 4), and the student's rank in future course grades (column 5). In all columns except column 5, one observation is a student; in column 5, one observation is a student-course pair. *Proximity* is the average frontier knowledge proximity of the courses taken by each student, measured in course-level standard deviations; in column 5, it is calculated using only courses taken before the focal course. All specifications control for average lagged proximity, instructor fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses in all columns except column 5, where standard errors are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A7: Educational Effects of Frontier Knowledge Exposure on Future Grades: Placebo Test

	Undergraduate	Graduate
	(1)	(2)
Future proximity (sd)	-0.0004	-0.0026
	(0.0012)	(0.0023)
Mean dep. var.	0.531	0.531
R^2	0.084	0.140
N	3,365,146	163,882
N clusters (students)	147,758	22,538

Notes: OLS estimates; one observation is a student-course pair. The dependent variable is the student's grade rank in the course. The variable *Future proximity* is the average frontier knowledge proximity of all courses taken by each student after the focal course. All specifications control for average lagged future proximity, instructor and school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Column 1 shows estimates for undergraduate students and column 2 shows estimates for graduate students. Standard errors in parentheses are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A8: Educational Effects of Frontier Knowledge Exposure - No Instructor Fixed Effects

Panel (a) Undergraduate students					
	Graduates	Time-to degree	GPA	Attends grad school	Future grade rank
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.063***	-0.167***	0.095***	0.011***	0.007***
	(0.004)	(0.017)	(0.009)	(0.003)	(0.001)
Mean dep. var.	0.814	5.724	2.822	0.123	0.524
R^2	0.230	0.354	0.225	0.185	0.060
N	189,163	154,101	188,166	189,163	3,671,479

Panel (b) Graduate students						
	MA			PhD		
	Graduates	Time-to degree	GPA	Graduates	Time-to degree	GPA
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.027***	-0.155***	0.028**	-0.008	-0.156**	0.041**
	(0.008)	(0.025)	(0.011)	(0.011)	(0.078)	(0.019)
Mean dep. var.	0.719	2.215	3.334	0.345	6.145	2.853
R^2	0.334	0.488	0.332	0.565	0.532	0.655
N	27,949	23,456	27,752	6,093	2,290	6,069

Notes: OLS estimates. Panel (a) reports results for undergraduate students and panel (b) for graduate students. The dependent variable is degree completion (column 1 in panel (a) and columns 1 and 4 in panel (b)), time-to-degree in years (column 2 in panel (a) and columns 2 and 5 in panel (b)), GPA (column 3 in panel (a) and columns 3 and 6 in panel (b)), graduate school attendance within Texas public universities (column 4 in panel (a)), and the student's rank in future course grades (column 5 in panel (a)). In all columns except column 5 of panel (a), one observation is a student; in column 5 of panel (a), one observation is a student-course pair. *Proximity* is the average frontier knowledge proximity of the courses taken by each student, measured in course-level standard deviations; in column 5 of panel (a), it is calculated using only courses taken before the focal course. All specifications control for average lagged proximity, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses in all columns except column 5 of panel (a), where standard errors are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A9: Earnings Effects of Frontier Knowledge Exposure - No Instructor Fixed Effects

Panel (a) Undergraduate students					
	ln(earnings)			Prob. earnings in...	
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.022***	0.037***	0.013	-0.029***	-0.002
	(0.007)	(0.008)	(0.010)	(0.004)	(0.003)
R^2	0.193	0.182	0.195	0.111	0.223
N	189,163	157,018	86,141	189,163	189,163
Panel (b) Graduate students					
	ln(earnings)			Prob. earnings in...	
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.043***	0.052***	0.019	-0.019***	0.020***
	(0.011)	(0.013)	(0.017)	(0.006)	(0.007)
R^2	0.364	0.372	0.364	0.258	0.390
N	32,413	26,825	15,494	32,413	32,413

Notes: OLS estimates; one observation is a student. The dependent variable is the natural logarithm of earnings 1-6 years (column 1), 1-3 years (column 2), and 4-6 years after predicted graduation (column 3); and an indicator for earnings in the bottom quartile (column 4) and the top quartile of the graduation cohort-specific earnings distribution (column 5). Panel (a) shows estimates for undergraduate students and panel (b) shows estimates for graduate students. The variable *Proximity* is the average frontier knowledge proximity of all courses taken by each student, measured in course-level standard deviations. All specifications control for average lagged proximity, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A10: Educational Effects of Frontier Knowledge Exposure - Excluding Major Course Overhauls

Panel (a) Undergraduate students					
	Graduates	Time-to degree	GPA	Attends grad school	Future grade rank
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.045***	-0.182***	0.110***	0.008	0.005***
	(0.006)	(0.025)	(0.013)	(0.005)	(0.002)
Mean dep. var.	0.872	5.669	2.884	0.117	0.524
R^2	0.558	0.593	0.448	0.284	0.085
N	189,917	154,614	188,922	189,917	3,670,228

Panel (b) Graduate students						
	MA			PhD		
	Graduates	Time-to degree	GPA	Graduates	Time-to degree	GPA
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.001	-0.108***	0.013	-0.024	0.345	0.089
	(0.013)	(0.041)	(0.019)	(0.031)	(0.476)	(0.057)
Mean dep. var.	0.713	2.293	3.354	0.259	6.295	2.808
R^2	0.512	0.703	0.515	0.823	0.893	0.797
N	25,903	21,523	25,715	4,561	1,215	4,539

Notes: OLS estimates. Panel (a) reports results for undergraduate students and panel (b) for graduate students. The dependent variable is degree completion (column 1 in panel (a) and columns 1 and 4 in panel (b)), time-to-degree in years (column 2 in panel (a) and columns 2 and 5 in panel (b)), GPA (column 3 in panel (a) and columns 3 and 6 in panel (b)), graduate school attendance within Texas public universities (column 4 in panel (a)), and the student's rank in future course grades (column 5 in panel (a)). In all columns except column 5 of panel (a), one observation is a student; in column 5 of panel (a), one observation is a student-course pair. *Proximity* is the average frontier knowledge proximity of the courses taken by each student, measured in course-level standard deviations and constructed excluding courses with an update of more than one standard deviation; in column 5 of panel (a), it is calculated using only courses taken before the focal course. All specifications control for average lagged proximity, instructor fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses in all columns except column 5 of panel (a), where standard errors are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A11: Earnings Effects of Frontier Knowledge Exposure - Excluding Major Course Overhauls

Panel (a) Undergraduate students						
	ln(earnings)			Prob. earnings in...		ln(industry
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile	earnings)
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.016	0.027**	0.012	-0.015**	0.006	0.003
	(0.011)	(0.013)	(0.018)	(0.007)	(0.006)	(0.007)
R^2	0.291	0.295	0.348	0.210	0.310	0.271
N	189,913	157,799	85,982	189,913	189,913	189,913
Panel (b) Graduate students						
	ln(earnings)			Prob. earnings in...		ln(industry
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile	earnings)
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.048**	0.070***	0.040	-0.029***	0.006	0.015
	(0.020)	(0.024)	(0.036)	(0.010)	(0.012)	(0.010)
R^2	0.512	0.537	0.566	0.452	0.518	0.561
N	30,197	24,755	13,345	30,197	30,197	30,197

Notes: OLS estimates; one observation is a student. The dependent variable is the natural logarithm of earnings 1-6 years (column 1), 1-3 years (column 2), and 4-6 years after predicted graduation (column 3); an indicator for earnings in the bottom quartile (column 4) and the top quartile of the graduation cohort-specific earnings distribution (column 5); and the natural logarithm of mean earnings in the first industry of employment post-graduation (column 6). Panel (a) shows estimates for undergraduate students and panel (b) shows estimates for graduate students. The variable *Proximity* is the average frontier knowledge proximity of all courses taken by each student, measured in course-level standard deviations and constructed excluding courses with an update of more than one standard deviation. All specifications control for average lagged proximity, instructor and school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A12: Educational Effects of Frontier Knowledge Exposure, Excluding Students Who Drop Courses

Panel (a) Undergraduate students					
	Graduates	Time-to degree	GPA	Attends grad school	Future grade rank
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.047***	-0.208***	0.092***	0.013***	0.003**
	(0.005)	(0.022)	(0.011)	(0.004)	(0.002)
Mean dep. var.	0.893	5.614	2.998	0.128	0.540
R^2	0.502	0.545	0.463	0.283	0.089
N	184,593	153,003	184,349	184,593	2,981,627

Panel (b) Graduate students						
	MA			PhD		
	Graduates	Time-to degree	GPA	Graduates	Time-to degree	GPA
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.012	-0.135***	0.041**	0.007	0.085	0.056
	(0.012)	(0.037)	(0.017)	(0.020)	(0.464)	(0.047)
Mean dep. var.	0.725	2.307	3.390	0.255	6.307	2.818
R^2	0.506	0.693	0.526	0.820	0.882	0.789
N	25,802	21,835	25,752	4,824	1,269	4,804

Notes: OLS estimates on the sample of students who do not drop any course in a given academic year, thereby excluding the students most likely to have added courses after observing the updated syllabus. Panel (a) reports results for undergraduate students and panel (b) for graduate students. The dependent variable is degree completion (column 1 in panel (a) and columns 1 and 4 in panel (b)), time-to-degree in years (column 2 in panel (a) and columns 2 and 5 in panel (b)), GPA (column 3 in panel (a) and columns 3 and 6 in panel (b)), graduate school attendance within Texas public universities (column 4 in panel (a)), and the student's rank in future course grades (column 5 in panel (a)). In all columns except column 5 of panel (a), one observation is a student; in column 5 of panel (a), one observation is a student-course pair. *Proximity* is the average frontier knowledge proximity of the courses taken by each student, measured in course-level standard deviations; in column 5 of panel (a), it is calculated using only courses taken before the focal course. All specifications control for average lagged proximity, instructor fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses in all columns except column 5 of panel (a), where standard errors are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A13: Earnings Effects of Frontier Knowledge Exposure - Excluding Students Who Drop Courses

Panel (a) Undergraduate students					
	ln(earnings)			Prob. earnings in...	
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.025**	0.038***	0.020	-0.027***	0.005
	(0.010)	(0.011)	(0.016)	(0.006)	(0.005)
R^2	0.291	0.297	0.349	0.211	0.313
N	184,595	153,256	83,790	184,595	184,595
Panel (b) Graduate students					
	ln(earnings)			Prob. earnings in...	
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.050***	0.075***	0.014	-0.018*	0.017
	(0.018)	(0.021)	(0.033)	(0.009)	(0.010)
R^2	0.512	0.537	0.563	0.455	0.517
N	30,389	24,894	13,498	30,389	30,389

Notes: OLS estimates on the sample of students who do not drop any course in a given academic year, thereby excluding the students most likely to have added courses after observing the updated syllabus; one observation is a student. The dependent variable is the natural logarithm of earnings 1-6 years (column 1), 1-3 years (column 2), and 4-6 years after predicted graduation (column 3); and an indicator for earnings in the bottom quartile (column 4) and the top quartile of the graduation cohort-specific earnings distribution (column 5). Panel (a) shows estimates for undergraduate students and panel (b) shows estimates for graduate students. The variable *Proximity* is the average frontier knowledge proximity of all courses taken by each student, measured in course-level standard deviations. All specifications control for average lagged proximity, instructor and school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A14: Educational Effects of Frontier Knowledge Exposure - Controlling for Syllabi Grading Structure

Panel (a) Undergraduate students					
	Graduates	Time-to degree	GPA	Attends grad school	Future grade rank
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.026***	-0.242***	0.065***	0.009**	0.003**
	(0.005)	(0.023)	(0.013)	(0.005)	(0.001)
Mean dep. var.	0.873	5.669	2.885	0.116	0.524
R^2	0.564	0.598	0.455	0.287	0.085
N	190,409	155,004	189,405	190,409	3,671,035

Panel (b) Graduate students						
	MA			PhD		
	Graduates	Time-to degree	GPA	Graduates	Time-to degree	GPA
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.016	-0.146***	0.026	-0.002	0.129	0.068
	(0.012)	(0.039)	(0.018)	(0.024)	(0.543)	(0.048)
Mean dep. var.	0.714	2.294	3.355	0.257	6.302	2.810
R^2	0.511	0.701	0.519	0.822	0.894	0.791
N	26,281	21,859	26,092	4,837	1,266	4,815

Notes: OLS estimates. Panel (a) reports results for undergraduate students and panel (b) for graduate students. The dependent variable is degree completion (column 1 in panel (a) and columns 1 and 4 in panel (b)), time-to-degree in years (column 2 in panel (a) and columns 2 and 5 in panel (b)), GPA (column 3 in panel (a) and columns 3 and 6 in panel (b)), graduate school attendance within Texas public universities (column 4 in panel (a)), and the student's rank in future course grades (column 5 in panel (a)). In all columns except column 5 of panel (a), one observation is a student; in column 5 of panel (a), one observation is a student-course pair. *Proximity* is the average frontier knowledge proximity of the courses taken by each student, measured in course-level standard deviations; in column 5 of panel (a), it is calculated using only courses taken before the focal course. In addition to the baseline controls (average lagged proximity, instructor fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile), all specifications include controls for courses' grading activities (i.e., average shares of terms in the assignments/grading portion of a syllabus that refer to exams, homework, and reports) as well as their lags. Robust standard errors are reported in parentheses in all columns except column 5 of panel (a), where standard errors are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A15: Earnings Effects of Frontier Knowledge Exposure - Controlling for Syllabi Grading Structure

Panel (a) Undergraduate students						
	ln(earnings)			Prob. earnings in...		ln(industry
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile	earnings)
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.022**	0.033***	0.015	-0.019***	0.009*	0.002
	(0.011)	(0.012)	(0.017)	(0.007)	(0.005)	(0.007)
R^2	0.292	0.297	0.351	0.211	0.311	0.272
N	190,405	158,199	86,227	190,405	190,405	190,405
Panel (b) Graduate students						
	ln(earnings)			Prob. earnings in...		ln(industry
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile	earnings)
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.051***	0.088***	0.022	-0.020**	0.019*	0.015
	(0.018)	(0.022)	(0.033)	(0.009)	(0.011)	(0.009)
R^2	0.512	0.538	0.561	0.454	0.515	0.560
N	30,882	25,316	13,693	30,882	30,882	30,882

Notes: OLS estimates; one observation is a student. The dependent variable is the natural logarithm of earnings 1-6 years (column 1), 1-3 years (column 2), and 4-6 years after predicted graduation (column 3); an indicator for earnings in the bottom quartile (column 4) and the top quartile of the graduation cohort-specific earnings distribution (column 5); and the natural logarithm of mean earnings in the first industry of employment post-graduation (column 6). Panel (a) shows estimates for undergraduate students and panel (b) shows estimates for graduate students. The variable *Proximity* is the average frontier knowledge proximity of all courses taken by each student, measured in course-level standard deviations. In addition to the baseline controls (average lagged proximity, instructor fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile), all specifications include controls for courses' grading activities (i.e., average shares of terms in the assignments/grading portion of a syllabus that refer to exams, homework, and reports) as well as their lags. Robust standard errors are reported in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A16: Educational Effects of Frontier Knowledge Exposure - Controlling for Instructor Experience

Panel (a) Undergraduate students					
	Graduates	Time-to degree	GPA	Attends grad school	Future grade rank
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.042***	-0.206***	0.108***	0.012***	0.003*
	(0.005)	(0.023)	(0.013)	(0.005)	(0.001)
Mean dep. var.	0.873	5.669	2.885	0.116	0.524
R^2	0.561	0.598	0.451	0.287	0.087
N	190,443	155,022	189,437	190,443	3,673,834
Panel (b) Graduate students					
	MA			PhD	
	Graduates	Time-to degree	GPA	Graduates	GPA
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.012	-0.128***	0.026	0.010	0.055*
	(0.012)	(0.038)	(0.018)	(0.017)	(0.031)
Mean dep. var.	0.713	2.329	3.362	0.252	2.812
R^2	0.511	0.699	0.510	0.817	0.788
N	26,428	21,977	26,231	4,850	4,828

Notes: OLS estimates. Panel (a) reports results for undergraduate students and panel (b) for graduate students. The dependent variable is degree completion (column 1 in panel (a) and columns 1 and 4 in panel (b)), time-to-degree in years (column 2 in panel (a) and column 2 in panel (b)), GPA (column 3 in panel (a) and columns 3 and 5 in panel (b)), graduate school attendance within Texas public universities (column 4 in panel (a)), and the student's rank in future course grades (column 5 in panel (a)). In all columns except column 5 of panel (a), one observation is a student; in column 5 of panel (a), one observation is a student-course pair. *Proximity* is the average frontier knowledge proximity of the courses taken by each student, measured in course-level standard deviations; in column 5 of panel (a), it is calculated using only courses taken before the focal course. All specifications control for average lagged proximity, instructor fixed effects, instructor experience fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses in all columns except column 5 of panel (a), where standard errors are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A17: Earnings Effects of Frontier Knowledge Exposure - Controlling for Instructor Experience

Panel (a) Undergraduate students					
	ln(earnings)			Prob. earnings in...	
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.023**	0.034***	0.015	-0.020***	0.008
	(0.010)	(0.012)	(0.017)	(0.006)	(0.005)
R^2	0.291	0.295	0.349	0.211	0.309
N	190,443	158,230	86,223	190,443	190,443
Panel (b) Graduate students					
	ln(earnings)			Prob. earnings in...	
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.041**	0.060***	0.026	-0.015	0.013
	(0.017)	(0.021)	(0.032)	(0.009)	(0.010)
R^2	0.508	0.534	0.556	0.450	0.511
N	31,004	25,405	13,779	31,004	31,004

Notes: OLS estimates; one observation is a student. The dependent variable is the natural logarithm of earnings 1-6 years (column 1), 1-3 years (column 2), and 4-6 years after predicted graduation (column 3); and an indicator for earnings in the bottom quartile (column 4) and the top quartile of the graduation cohort-specific earnings distribution (column 5). Panel (a) shows estimates for undergraduate students and panel (b) shows estimates for graduate students. The variable *Proximity* is the average frontier knowledge proximity of all courses taken by each student, measured in course-level standard deviations. All specifications control for average lagged proximity, instructor fixed effects, instructor experience fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Robust standard errors are reported in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A18: Educational Effects of Frontier Knowledge Exposure - Controlling for Instructor Teaching Load

Panel (a) Undergraduate students					
	Graduates	Time-to degree	GPA	Attends grad school	Future grade rank
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.042***	-0.214***	0.108***	0.012***	0.004***
	(0.005)	(0.023)	(0.013)	(0.005)	(0.001)
Mean dep. var.	0.873	5.669	2.885	0.116	0.524
R^2	0.561	0.596	0.451	0.287	0.085
N	190,443	155,022	189,437	190,443	3,673,834
Panel (b) Graduate students					
	MA			PhD	
	Graduates	Time-to degree	GPA	Graduates	GPA
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.017	-0.141***	0.021	0.014	0.032
	(0.017)	(0.054)	(0.025)	(0.024)	(0.064)
Mean dep. var.	0.713	2.321	3.361	0.253	2.811
R^2	0.589	0.765	0.595	0.864	0.805
N	24,265	19,904	24,064	3,805	3,785

Notes: OLS estimates. Panel (a) reports results for undergraduate students and panel (b) for graduate students. The dependent variable is degree completion (column 1 in panel (a) and columns 1 and 4 in panel (b)), time-to-degree in years (column 2 in panel (a) and column 2 in panel (b)), GPA (column 3 in panel (a) and columns 3 and 5 in panel (b)), graduate school attendance within Texas public universities (column 4 in panel (a)), and the student's rank in future course grades (column 5 in panel (a)). In all columns except column 5 of panel (a), one observation is a student; in column 5 of panel (a), one observation is a student-course pair. *Proximity* is the average frontier knowledge proximity of the courses taken by each student, measured in course-level standard deviations; in column 5 of panel (a), it is calculated using only courses taken before the focal course. In addition to the baseline controls (average lagged proximity, instructor fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile), all specifications include fixed effects for instructors' teaching load (number of sections taught in a given year). Robust standard errors are reported in parentheses in all columns except column 5 of panel (a), where standard errors are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table A19: Earnings Effects of Frontier Knowledge Exposure - Controlling for Instructor Teaching Load

Panel (a) Undergraduate students					
	ln(earnings)			Prob. earnings in...	
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.024**	0.035***	0.017	-0.020***	0.008
	(0.010)	(0.012)	(0.016)	(0.006)	(0.005)
R^2	0.292	0.297	0.350	0.211	0.311
N	190,443	158,230	86,223	190,443	190,443
Panel (b) Graduate students					
	ln(earnings)			Prob. earnings in...	
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.062**	0.077**	0.033	-0.029**	0.020
	(0.025)	(0.030)	(0.049)	(0.013)	(0.015)
R^2	0.594	0.629	0.614	0.556	0.592
N	28,803	23,146	12,106	28,803	28,803

Notes: OLS estimates; one observation is a student. The dependent variable is the natural logarithm of earnings 1-6 years (column 1), 1-3 years (column 2), and 4-6 years after predicted graduation (column 3); and an indicator for earnings in the bottom quartile (column 4) and the top quartile of the graduation cohort-specific earnings distribution (column 5). Panel (a) shows estimates for undergraduate students and panel (b) shows estimates for graduate students. The variable *Proximity* is the average frontier knowledge proximity of all courses taken by each student, measured in course-level standard deviations. In addition to the baseline controls (average lagged proximity, instructor fixed effects, school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile), all specifications include fixed effects for instructors' teaching load (number of sections taught in a given year). Robust standard errors are reported in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Appendix B Dataset Construction and Frontier Knowledge Measurement

We gathered our sample of syllabi from the websites of seven public universities in Texas. This appendix describes how we construct the syllabus corpus, extract academic content from each document, link syllabi to course metadata, and combine syllabus text with academic publication data to construct frontier knowledge proximity.

The construction proceeds in four steps. We first collect and clean syllabi from university repositories. We then identify the academic-content sections of each syllabus and extract course metadata. Next, we construct a publication corpus from OpenAlex and process both syllabi and publications into comparable term-frequency vectors. Finally, we use these vectors to compute each syllabus's proximity to the contemporary research frontier and validate this measure.

B.1 Texas Syllabus Corpus

Our syllabi sample covers the majority of courses taught at seven major public universities in Texas: Stephen F. Austin State University (starting from 2009), Sam Houston State University (2011), Texas A&M University (2013), University of Houston-Clear Lake (2010), University of Texas at Austin (2011), University of Texas at Dallas (2005), and West Texas A&M University (2013). We collected these syllabi directly from each university's website. These universities make these documents available for download (see Appendix Table B1), and we downloaded them between October 2023 and February 2024.[17] Our sample includes a total of 459,415 documents corresponding to 27,872 courses taught up to 2022. These account for approximately 51% of all courses offered at these institutions during our analysis period. Coverage varies across institutions and years depending on the availability of historical online repositories.

B.2 Cleaning Syllabus Text and Extracting Academic Content

To transform raw syllabus files into usable academic content, we (i) extract text from PDF files; (ii) clean the extracted text; (iii) identify syllabus sections; and (iv) retain sections that describe course content, readings, objectives, requirements, and evaluation while removing administrative or policy material unrelated to academic content. We describe these steps below.

B.2.1 Cleaning Raw Syllabus Text

To clean the raw text of each syllabus, we proceed as follows:

[17] We contacted all public universities in Texas that do not make historical syllabi available online to request access to their records. However, most universities were unable to provide these documents because Texas House Bill 2504 of 2009 only requires public colleges and universities to maintain records for two years following each syllabus' term of instruction.

Table B1: Texas Syllabi: General Information

Institution	Years available	Link
Sam Houston State U	2011-23	https://samweb.shsu.edu/faci10wp/
Stephen F. Austin State U	2009-24	https://orion.sfasu.edu/courseinformation/
Texas A&M U	2013-23	https://howdy.tamu.edu/uPortal/p/public-class-search-ui.ctf1/max/render.uP#
U of Texas at Austin	2011-23	https://utdirect.utexas.edu/apps/student/coursedocs/nlogon/?year=&semester=&department=GEO&course_number=420K&course_title=&unique=&instructor_first=&instructor_last=&course_type=In+Residence&search=Search
U of Texas at Dallas	2005-24	https://coursebook.utdallas.edu/
U of Houston–Clear Lake	2010-24	https://saprd.my.uh.edu/psc/saprd/EMPLOYEE/HRMS/c/UHS_SS_CUSTOM.UHS_HB2504_DISPLAY.GBL?institution_name=UHCL&
West Texas A&M U	2013-24	https://syllabus.wtamu.edu/syllabi/

Note: List of public universities in Texas included in our sample, with years of syllabi availability and links to webpages containing the syllabi. We downloaded syllabi in October 2023-February 2024.

(i) We first convert all PDF files of syllabi to plain text using PyMuPDF.[18]

(ii) We use the Unidecode Python package[19] to convert Unicode text into ASCII text. This allows us to handle legacy code that does not support Unicode, non-Roman names on a US keyboard, and ASCII approximations for symbols and non-Latin alphabets.

(iii) We remove browser information, often present in the header of a syllabus, by searching for keywords such as "Internet Explorer", "Newer Browser", "JavaScript Enabled", "Cookies Are", "Download Info", "Login", "Log In", "Print", and "Search".

B.2.2 Identifying Academic Content Within Syllabi

Most syllabi are organized into sections, only some of which are relevant for measuring course content. The relevant sections include instructor and course information, course descriptions, course re-

[18] https://pymupdf.readthedocs.io/en/latest/
[19] https://pypi.org/project/Unidecode/

quirements, learning objectives, outlines, readings, homework, exams, and other evaluation methods. Syllabi also often include administrative or policy material that is not used in the construction of the frontier knowledge measure, such as honor-code statements, disability-related policies, classroom laptop and cellphone policies, attendance rules, and other course policies.

To parse sections, we developed a supervised algorithm using a set of section-title keywords. The algorithm identifies a section type by searching for keywords associated with each category. Table B2 provides section types along with the corresponding keywords. Using these keywords, the algorithm partitions the text into sections of the syllabus by combining keyword matches with each syllabus's formatting rules. Figure B1 illustrates the process using part of a syllabus.

1. For each syllabus, we identify section titles based on the word list described above and on formatting features. We mark all cases in which section-title phrases appear as all uppercase text or as consecutive initial capital letters using regular expressions. In Figure B1, underlined phrases satisfy the features of a section title, such as "Course Description."

2. We divide the syllabus into parts and use Arabic numerals to mark them. We then select sections with relevant titles and extract the cleaned academic-content text.

3. In Figure B1, we focus on highlighted sections, such as "Course Objective," "Prerequisites," and "Text."

B.3 Course Metadata

B.3.1 Course Codes

We use course codes and titles to identify repeated offerings of the same course within an institution and to group similar courses across institutions. This step is necessary because course numbering systems are institution-specific and may change over time.

Our data extraction process allows us to obtain the course code corresponding to each syllabus. However, these codes are institution-specific and often vary over time. To identify courses of the same level (e.g., basic undergraduate) covering the same topic (e.g., Principles of Microeconomics), both within and across schools, we proceed in two steps.

First, we construct a unified within-school course code using the raw course code and the course name. We remove punctuation and multiple whitespaces from codes and names. For course names, we also remove stop words and isolate the course stem name, i.e., the common base form of the

Table B2: Section Title Keywords List

Section type	Keywords
Course Description	Syllabi, Syllabus, Title, Description, Method, Instruction, Content, Characteristics, Overview, Tutorial, Intro, Abstract, Methodologies, Summary, Conclusion, Appendix, Guide, Document, Module, Introduction, Approach, Lab, Background
Requirements	Requirement, Applicability, Required
Objectives	Objectives, Achievement, Outcome, Motivation, Purpose, Statement, Skill, Competency, Performance, Goal
Outline	Outline, Schedule, Timeline, Guideline
Materials	Text, Material, Resource, Recommend, Reference, Book, Calendar, Textbook, Guidebook
Instructor information	Instructor, About, Email, Phone, Contact, Professor, Staff, Faculty, Information
Projects, homework, papers, and exams	Personal, Total, Individual, Exercise, Essay, Submission, Assign, Homework, Paper, Final, Examining, Midterm, Term, Semester, Proposal, Application, Demonstration, Program, Task, Report, Practical, Drafting, Project, Plan, Deadline, Makeup, Advising, Advisor, Survey, Assignment, Planning, Practice, Group, Participation, Team, Research, Activity, Complaint, Design, Analysis, Strategy, Procedure, Working, Work, Exam, Examination, Training, Professional, Test, Case, Discussion, Grade, Presentation, Quiz, Essay, Layout, Sample, Rewrite
Grades	Assessment, Point, Scope, Evaluation, Record, Grading, Composition, Review
Other Policies	Academic, Justice, Administration, Rule, Discipline, Disclaimer, Regulation, Standard, Affair, Dishonesty, Plagiarism, Misconduct, Offence, Medical, Absent, Absence, Trip, Religious, Observance, Attendance, Honesty, Origination, Originator, Help, Technology, Attendance, Accessing, Service, Opportunity, Administrative, Accommodation, Support, Policy, Right, Responsibility, Disability, Weather, Integrity, Copyright
Notes	Remark, Notice, Additional, Acknowledgement, Absolutely, Absolute, Important, Note, Cannot, Can, Must, Should, Will, Please, No
Other Words	Course, Lecture, Catalog, Campus, Community, Class, Classroom, College, University, Discussion, Seminar

Note: Keywords used to identify the corresponding section types of a syllabus. In the implementation, we use both the singular and plural versions of each term.

words. We then consider two courses as sharing a course code if either (a) they share the same name and code or (b) they share the same name, even if the course code changes over time. This procedure accounts for the possibility that course-code systems change within a school over time.

Once we have a disambiguated identifier for courses within the same school, we assign courses a cross-school identifier. Specifically, we assign two courses the same cross-school identifier if they share the same standardized course name.

B.3.2 Course Level: Basic, Advanced, Graduate

We classify each syllabus into one of three levels—basic undergraduate, advanced undergraduate, or graduate—using a deterministic crosswalk between the course code and each university's published numbering convention. For most schools the leading digit of the course code identifies the level. The University of Texas at Austin is the exception: its leading digit denotes the course's semester credit-hour value, so the *second* digit carries the rank information.[20] Table B3 summarizes the rules.

Table B3: Course-level assignment rules by institution

Institution	Digit used	Basic	Advanced	Graduate
WTAMU, UHCL, UTDallas, SHSU[a]	1st	0–2	3–4	5–8
SFA [b]	1st	0–2	3–4	5–6
Texas A&M (TAMU)[c]	1st	1–2	3–4	5–9
UT Austin[d]	2nd	0–1	2–7	8–9

[a] Follows the Texas Common Course Numbering System (TCCNS, https://www.tccns.org/about/). [b] Per the SFA registrar's course numbering key: 1000-level freshman, 2000 sophomore, 3000 junior, 4000 senior, 5000 graduate, 6000 doctoral; courses prior to fall 2020 used a three-digit form with the same leading-digit semantics (https://www.sfasu.edu/registrar/transcripts/key). [c] Per the TAMU catalog: 100–199 freshman, 200–299 sophomore, 300–399 junior, 400–499 senior, with 500 and above graduate (https://catalog.tamu.edu/undergraduate/appendices/texas-common-course-numbering-system/). [d] The first digit gives credit-hour value; the last two digits give rank, with 01–19 lower-division, 20–79 upper-division, and 80–99 graduate.

These rules assign a level to 99.99% of the syllabi in our sample; the residual fraction with non-conforming codes is left missing. We adopt this rule-based classification in place of an earlier undergraduate / graduate flag carried in the raw metadata, which we found to misclassify non-

[20] https://extension.utexas.edu/credits-grades-and-transcripts.

trivial numbers of courses (for example, several 5000-level graduate courses at SHSU were tagged as undergraduate in the source data).

B.3.3 Instructor Names

To extract the name of the instructor from each syllabus, we build a neural network model based on the BiLSTM-CNNs-CRF model for named entity recognition (NER).[21] The training/test dataset is built via the following three steps:

(i) We select syllabi that contain at least one keyword such as "Doctor", "Doctors", "Dr", "Professor", "Prof", "Instructor", "Instructors", "Tutor", or "Tutors" in the first 3,500 characters.

(ii) We use the Spacy[22] package to identify whether the words following those keywords are names of people, i.e., whether the entity label is "PERSON."

(iii) We process the syllabus text sentence by sentence as the training and test data of the model.

We also apply a few additional filters: (a) we remove single-letter names; (b) we require all words in the name to appear in the Python Library *English First and Last Names Data Set*[23]; and (c) after applying the first two filters, we keep only the first instructor name. With this algorithm, we are able to assign an instructor name to 86.23% of all syllabi. The out-of-sample precision of this algorithm is 85.18%.

B.3.4 References and Recommended Readings in Each Syllabus

To extract assigned readings and course materials from each syllabus, we employ a large language model (LLM)-based extraction pipeline. Specifically, we use OpenAI's GPT-4.1 mini model[24] to identify and extract all course materials mentioned in each syllabus. This approach allows us to systematically identify the readings and materials assigned in each course, which we subsequently use to analyze patterns in course content and to validate the frontier knowledge proximity measure. Next, we use regular expressions to extract publication years, when available, from the extracted reference strings. In our sample, we successfully extract reference information for 80.4% of non-empty syllabi and obtain at least one publication year for 42.0% of non-empty syllabi; among syllabi with extracted years, we obtain an average of 1.77 publication years per syllabus. The prompt used for extraction is shown below:

[21] BiLSTM-CNNs-CRF model for named entity recognition (NER), Ma and Hovy (2016).
[22] https://spacy.io/
[23] https://github.com/philipperemy/name-dataset
[24] https://platform.openai.com/docs/models/gpt-4.1-mini

Prompt for Reference Extraction

You will be given a syllabus text. Your task is to extract reference information from this text without making any changes to the content. To extract the reference information, follow these steps:

1. *Look for sections in the syllabus that contain information about course materials, textbooks, or references. Common prompt words to look for include: Materials, Text, Material, Resource, Recommend, Reference, Book, Calendar, Textbook, Guidebook.*
2. *Identify sections that list specific items like books, articles, or academic papers.*
3. *Extract only those parts listing the materials—omit any sentences about how or when to access, read, or bring the materials.*
4. *If multiple relevant sections are found, extract all and merge them into one for the output, separating each source with "|".*
5. *If you don't find any relevant sections containing reference information, respond with "None".*

Output Generation: Output only the TSV data without any additional text or formatting. Include columns: Reference. Ensure proper TSV formatting with no extra line breaks or spaces. Use "None" to represent missing values.

B.4 Academic Publication Corpus

To construct our measure of frontier knowledge proximity, we collect a large sample of research outputs from OpenAlex. OpenAlex is an open scholarly knowledge graph maintained by OurResearch that provides large-scale metadata and linkage information on research outputs and related entities. It covers works, including journal articles, conference papers, and preprints, as well as authors, institutions, sources, topics, publishers, and funders, and encodes relationships such as authorship, affiliations, citations, and topical classifications. OpenAlex is openly licensed under CC0 and is accessible via a public API and periodic data snapshots.

B.4.1 Collecting Academic Publications

We use the following variables:[25]

- `id`: OpenAlex ID, used as the unique identifier of each article;
- `title`: title of the article;

[25] https://openalex.org. The full list of variables available through OpenAlex is documented at https://docs.openalex.org/api-entities/works/work-object.

- `issn`: ISSN of the journal;
- `publication_date`: publication date;
- `abstract_inverted_index`: abstract, stored in inverted-index format;
- `keywords`: author-provided keywords.

From an initial pool of 261,381,162 OpenAlex records, we identified 133,379,066 valid entries and ultimately retained 107,911,700 English-language articles with non-empty abstracts for our analysis.

B.4.2 Cleaning Publication Abstracts

The main information from academic publications that we use in the analysis is the abstract. In OpenAlex, abstracts are stored in the `abstract_inverted_index` field as an inverted index, where each unique word is mapped to the positions in which it appears in the text. We first reconstruct the full abstract text by reversing this inverted-index structure.[26]

We further clean this variable to remove copyright disclaimers, usually present at the beginning or end of each abstract and unrelated to content. We do this using keyword recognition techniques. Starting from the first sentence of an abstract, we remove it if it contains at least one of the following words: "copyright", "©", "published", "publisher", "all right", or "all rights reserved." We repeat this procedure until the first sentence does not contain any of these words. We apply the same procedure from the end of the abstract to remove trailing copyright or publisher statements.

B.5 Constructing Frontier Knowledge Proximity

We now describe how we transform cleaned syllabus text and cleaned publication abstracts into the frontier knowledge proximity measure used in the analysis.

B.5.1 Text Pre-processing

The preceding steps produce cleaned syllabus text and cleaned publication abstracts. Before constructing term-frequency vectors, we apply a common preprocessing procedure to both corpora.

First, we convert the text of each document into ASCII text using the Unidecode Python package.[27] Next, we convert all capitalized characters to lowercase and use the NLTK Python Toolkit to

[26] For example, an inverted index {"Purpose": [0], "of": [1], "study": [2]} is converted to "Purpose of study."

[27] https://pypi.org/project/Unidecode/

strip out non-word text elements, such as punctuation marks, numbers, and HTML tags. We also remove all occurrences of 280 stop words, which include prepositions, punctuation marks, pronouns, and other words that carry little semantic content.[28]

B.5.2 Dictionary Construction

The core of our baseline measure relies on a dictionary of academic knowledge terms. We construct this dictionary as the list of all unique terms used as keywords in the OpenAlex academic publication sample described in Section B.4. This restriction focuses the term space on vocabulary used to describe academic knowledge rather than on generic words that appear in syllabi for administrative or instructional reasons. We also construct alternative versions of the measure using broader n-gram representations of syllabus and publication text. These alternatives relax reliance on the keyword dictionary and allow the measure to capture frontier concepts expressed as short phrases.

Academic Publications Dictionary The first dictionary is the list of all unique words and expressions used as keywords in academic publications. We extract these keywords from the OpenAlex data described in Section B.4.

Syllabi-Based Empirical Dictionary The second dictionary is constructed empirically from the syllabus corpus itself. This data-driven approach ensures that our terminology reflects the pedagogical vocabulary used in higher education, rather than solely research-oriented jargon. At the same time, because the measure compares these terms to publication corpora over time, the frontier component continues to come from the relative alignment of syllabus language with recent versus older academic research.

To construct this dictionary, we begin with the cleaned text described in Section B.5.1 and apply two additional refinement steps:

1. **Context-Specific Filtering:** We remove dates, timestamps, and a curated list of syllabus-specific stop words, including administrative terms such as "office hours" or "prerequisites," to reduce noise.

2. **Frequency Thresholding:** We extract all unique N-grams ($N \in [1, 3]$) and impose a minimum document-frequency threshold. Only terms appearing in at least 100 distinct syllabi are retained. This threshold filters out idiosyncratic terms and preserves generalizable knowledge concepts.

[28] We create a list of stop words as the union of all single letters and the Stanford CoreNLP package: https://github.com/stanfordnlp/CoreNLP/blob/master/data/edu/stanford/nlp/patterns/surface/stopwords.txt.

The resulting dictionary contains 691,769 unique terms, approximately 50% of which intersect with the academic publications dictionary. The term-frequency distribution exhibits a characteristic heavy-tailed structure. For instance, while the dictionary is large, the "core" vocabulary is concentrated: only 9,204 terms appear in at least 1% of the syllabi, and 118,099 terms appear in at least 0.1%. Conversely, the vast majority of terms appear infrequently, reflecting the specialized nature of advanced coursework.

B.5.3 Term Extraction Algorithm

Once we have cleaned the text and established a dictionary, we convert the documents into numerical data using a term-extraction algorithm called NGramMatch. This algorithm performs exact string matching between the text of each document, represented as N-grams with N ranging from 1 to 3, and the dictionary. The algorithm extracts N-grams from each document to form a basic term set, filters out terms that cannot be linked to any dictionary entry, and assigns each document a frequency vector based on matched dictionary terms.

The output of this step is a sparse vector describing the knowledge terms that appear in each syllabus and each publication abstract. These vectors are the inputs into the frontier-proximity calculation described next.

B.5.4 Similarity to Recent and Older Publications

We use these term-frequency vectors to measure how close each syllabus is to the contemporary research frontier of its field. For each syllabus d taught in year $t(d)$, we calculate its average similarity with articles published in a three-year window centered τ years before the syllabus year:

$$S_d^\tau = \frac{\sum_{n \in \Omega_\tau(d)} \rho_{dn}}{|\Omega_\tau(d)|},$$

where $\Omega_\tau(d)$ includes all articles published in the period $[t(d) - \tau - 1, t(d) - \tau + 1]$, and ρ_{dn} is the cosine similarity between syllabus d and article n.
We then define frontier knowledge proximity as:

$$p_d \equiv 100 \times \left(\frac{S_d^\tau}{S_d^{\tau'}} \right), \tag{8}$$

where τ indexes the recent publication window and τ' indexes the older publication window, with $\tau' > \tau$. Higher values indicate that the syllabus is more similar to recent research than to older research.

We define τ and τ' separately by field, using the 5th and 90th percentiles of the distribution of citation ages within each field. This allows the definition of recent and older knowledge to vary with the pace of knowledge production across fields. The ratio form compares recent and older publication similarity within the same syllabus, which attenuates the influence of syllabus-specific features such as length, verbosity, and writing style that may raise or lower similarity to academic publications in general.

B.6 AI-Based Validation of Frontier Knowledge Proximity

One natural way to validate the measure would be to ask field specialists to assess whether syllabi reflect the current frontier of their disciplines. Because our corpus spans many fields and course levels, implementing such an exercise at scale would require a large set of specialized raters. We therefore complement our other validation exercises with an LLM-based content assessment. The goal of this exercise is to test whether our measure aligns with an independent assessment of the substantive knowledge covered in real syllabi.

To implement this validation, we examine whether a large language model (LLM), given only the substantive academic content of two otherwise-comparable syllabi, agrees with our measure on which of the two is closer to the knowledge frontier. Unlike the validation exercises reported in Section 3, this check relies neither on a syllabus's reference list nor on its nominal course level, and does not use simulated text. Instead, it asks whether a model-based assessment of the substantive concepts covered in the course, without access to the $TF\text{-}BIDF$ weighting scheme, to publication vintages, or to the syllabus-publication term-frequency vectors, ranks pairs of real syllabi in the same direction as our measure does. This exercise is designed to validate the substantive content captured by the measure: whether syllabi that score higher on frontier proximity also appear, when reduced to their core knowledge points, to cover more frontier-oriented material.

Pair construction and sampling. We start from the deduplicated Texas syllabus corpus described in Section B.1. Within each institution-semester-department prefix-course code cell and calendar year, we keep a single section, retaining the one with the longest cleaned text. We restrict attention to the three most recent academic years in the sample, require non-missing frontier knowledge proximity and a minimum cleaned-text length of 200 words, and drop syllabi whose cleaned text contains 5% or more non-printable or non-ASCII characters, typically the result of failed PDF extraction.

We then restrict the sample to broad STEM fields—Biology, Chemistry, Physics, Astronomy, Earth and Atmospheric Sciences, Computer Science, Engineering, Mathematics, Medicine, Veteri-

nary Medicine, Health Technician programs, and Agriculture. We focus on these fields because concepts, methods, and tools provide relatively clear markers of the knowledge frontier. Finally, we exclude non-lecture courses—seminars, colloquia, laboratories, independent or directed studies, internships, practica, theses, dissertations, reading courses, research courses, special or selected topics, capstones, clinics, workshops, field experiences, and tutorials—identified by regular-expression matches on the course title, course code, and body of the syllabus.

From the resulting pool, we construct all within-group pairs of distinct syllabi that share the same institution, course field, department prefix, course level, and calendar year. We drop pairs with the same course code, which correspond to different sections of the same course, and pairs taught by the same instructor. To ensure that paired syllabi are of comparable verbosity, we retain only pairs whose absolute log word-length ratio is at or below its median across all possible pairs. To ensure that each pair is informative about the measure, we retain only those whose absolute frontier-proximity gap is in the top quartile of its distribution across pairs. To ensure that paired syllabi cover overlapping subject matter, we compute the cosine similarity of the two documents' $TF\text{-}BIDF$ representations, using the same syllabus term-frequency matrix that underlies the main measure, and keep pairs whose cosine similarity lies between the 50th and 90th percentile of the filtered-pair distribution: pairs that are clearly on the same topic but not near-duplicates.

We then draw a stratified random sample of 50 pairs for AI evaluation, balanced across frontier-proximity levels, calendar years, and the direction of the within-pair proximity gap. No syllabus is allowed to appear in more than one sampled pair, so the 50 pairs involve 100 distinct syllabi.

LLM knowledge-point extraction. For each of the 100 sampled syllabi, we use an LLM, OpenAI's GPT-5 mini, to perform two tasks. First, the model anonymizes the syllabus by removing instructor and institution names, email addresses, phone numbers, office locations, and URLs, replacing them with generic placeholders. Second, it extracts an ordered list of the ten most important *subject-matter knowledge points* described in the syllabus.

Each knowledge point is constrained to be a concise noun phrase, two to eight words, naming a concept, theory, method, technique, tool, model, phenomenon, or topic from the discipline's body of knowledge—the kind of entry that would appear as a chapter heading or index entry in a leading textbook of the field. The prompt explicitly excludes generic teaching strategies, professional-development or career items, course logistics, assessment formats, generic transferable skills, and instructional-technology references, and requires that each knowledge point be sourced from the course description, learning objectives, weekly topic schedule, laboratory topics, lecture descrip-

tions, or the subject-matter content of the reading list, as opposed to bibliographic information about the readings themselves. Duplicates are removed and the list is capped at ten points per syllabus. The resulting object is therefore a compact representation of what the course teaches, rather than a summary of the age or identity of assigned readings.

Prompt for Knowledge-Point Extraction

You are an expert at analyzing academic course syllabi. Given a college course syllabus, perform the following tasks:

1. ***Anonymize.*** *Remove all instructor names, institution names, campus names, email addresses, phone numbers, office locations, and URLs from the text. Replace them with generic placeholders such as [INSTRUCTOR] or [INSTITUTION]. Keep the course subject, level, and year intact.*
2. ***Extract the top ten most important knowledge points.*** *Read the entire syllabus carefully and identify the ten most important subject-matter knowledge points that students are expected to learn. A knowledge point is a specific concept, theory, method, technique, tool, model, phenomenon, or topic from the academic discipline itself.*

The following rules apply when extracting knowledge points:

- *Each phrase should be a concise noun phrase (two to eight words), e.g., "Bayesian inference," "CRISPR gene editing," "supply and demand equilibrium," or "Monte Carlo simulation methods."*
- *Focus on the knowledge points that best represent the core intellectual content of the course—the central concepts, theories, and methods that define what the course is fundamentally about.*
- *Include knowledge points from the course description, learning objectives, weekly topic schedule, lab topics, lecture descriptions, reading assignments (extracting the subject-matter topics they cover), and any other substantive academic content.*
- *For reading materials (textbooks, articles, etc.), extract the subject-matter topics they cover, not the bibliographic information (title, author, year). For example, if a reading is "Smith (2020) Machine Learning for Healthcare," extract "machine learning for healthcare" as a knowledge point.*
- *Do not include pedagogical methods (e.g., "group discussion," "flipped classroom"), administrative items (e.g., "grading policy," "attendance"), assessment formats (e.g., "midterm exam," "final project"), or instructional technology (e.g., "Canvas LMS," "Zoom").*
- *Return exactly ten knowledge points. If the syllabus contains fewer than ten identifiable knowledge points, return as many as can be found.*
- *Remove duplicates—each unique knowledge point should appear only once.*

- *Order them by importance: the most central and essential knowledge point first.*

Output Generation: Return only a valid JSON object (no markdown fencing) with this exact structure:

```
{
  "knowledge_points":  ["phrase1", "phrase2", ..., "phrase10"],
  "n_knowledge_points":  <integer count>
}
```

If no knowledge points can be identified, return {"knowledge_points": [], "n_knowledge_points": 0}.

Pairwise LLM comparison. For each of the 50 sampled pairs, we ask the same model to compare the two ten-point knowledge-point lists and decide which syllabus covers more frontier, or cutting-edge, subject-matter knowledge in its field, based solely on the knowledge-point phrases. The prompt instructs the model to focus on whether one list contains more recent, cutting-edge, or state-of-the-art concepts, theories, methods, techniques, or tools, and to ignore the total number of knowledge points, the breadth of coverage of foundational topics, the course level, and any single buzzword. To guard against position bias, the two syllabi are presented in randomized A/B order across repetitions, and each pair is evaluated five times independently, yielding 250 comparisons in total. For each pair we take the majority vote across the five evaluations as the AI winner.

Because the comparison is based only on extracted knowledge points, the model does not observe the full syllabus text, the length or formatting of the document, the instructor, the institution, or the underlying frontier-proximity score.

Prompt for Pairwise Comparison

You are an expert at comparing academic course syllabi. You will receive two lists of top-ten subject-matter knowledge-point phrases extracted from two different syllabi. Your task is to determine which syllabus covers more frontier (cutting-edge) subject-matter knowledge in its field, based solely on the knowledge-point phrases.
Focus only on:

- *Whether the knowledge points include more recent, cutting-edge, or state-of-the-art concepts, theories, methods, techniques, or tools in the discipline.*
- *Whether the knowledge points reflect awareness of recent developments, discoveries, or paradigm shifts.*

- *Depth and specificity of advanced/frontier topics (e.g., "deep learning for natural language processing" is more frontier than "introduction to artificial intelligence").*

Ignore everything else, including:

- *The total number of knowledge points (more points do not imply more frontier).*
- *Breadth of coverage (covering more foundational topics does not make a syllabus more frontier).*
- *Pedagogical or administrative considerations (these have already been filtered out).*

Output Generation: Return only a valid JSON object with this exact structure:

```
{
  "winner": "A" or "B",
  "confidence": 1-10,
  "reason_tag": one of "newer_topics", "more_advanced_methods",
     "more_specialized", "mixed_signals", "unclear",
  "reasoning": "Brief 1-2 sentence explanation focusing on
     which knowledge points signal more frontier content"
}
```

If genuinely indistinguishable, set `winner` *to the one with even slightly more frontier signals.*

Concordance with the frontier measure. We then compare the AI winner of each pair with the syllabus whose frontier knowledge proximity is higher on our measure—the "expected winner." In 37 of the 50 pairs (74%) the AI's majority vote agrees with our measure. Under the null of random (i.e., 50%) agreement, this corresponds to a one-sided binomial test with $p = 0.00047$. This level of agreement is notable considering that the AI's judgment is based on only twenty short phrases per pair, ten per syllabus, with no access to the full term-frequency vector, to the backward-looking IDF weights, or to any information about when particular concepts entered the research frontier.

We interpret this exercise as complementary evidence on the construct validity of the measure. Among comparable STEM lecture courses, syllabi that score higher on our frontier knowledge proximity measure are also more likely to be judged, from their extracted knowledge points alone, as covering more frontier-oriented subject matter. The exercise does not rely on reference-list recency and is not driven by formatting, verbosity, or course logistics, since those features are removed before the comparison.

Figure B1: Dividing A Syllabus Into Sections: An Example

Econ 561a Yale University Fall 2005
Prof. Tony Smith (Part I) Prof. Michael Keane (Part II)
Syllabus for **COMPUTATIONAL METHODS FOR ECONOMIC DYNAMICS** ECON 561a

Course Objectives:
Most of the dynamic economic models used in modern quantitative research in economics do not have analytical (closed-form) solutions. For this reason, the computer has become an indispensable tool for conducting research in dynamic economics. The goal of this two-part course is precisely to teach students computational tools for conducting numerical analysis of dynamic economic models. The focus of the first half of the course, taught by Prof. Tony Smith, is on solving dynamic programming problems and on computing competitive equilibria of dynamic economic models. The first half of the course also provides an introduction to some of the basic tools of numerical analysis, including minimization, root-finding, interpolation, function approximation, and integration. The focus of the second half course, taught by Prof. Michael Keane, is on solving and estimating discrete-choice dynamic programming models of economic behavior. Taken together, the two halves of the course provide students with a thorough introduction to the numerical analysis of dynamic economic models in both microeconomics and macroeconomics.

Contact Information (Prof. Tony Smith)
Office: 28 Hillhouse, Room 306 Office phone: (203) 432-3583
Email address: tony.smith@yale.edu Course Web site: www.econ.yale.edu/smith/econ561a
Office hours: Thursdays from 10AM–noon, or by appointment

Class Meetings:
The course meets on Mondays and Wednesdays from 2:30PM to 3:50PM in a room to be determined.

Prerequisites:
This course is designed for graduate students in economics who have taken first-year graduate courses in microeconomics, macroeconomics, and econometrics. No prior knowledge of either numerical methods or computer programming is assumed, but some familiarity with a programming language would prove helpful.

Texts:
The required textbook for this course is:
Numerical Recipes in Fortran 77: The Art of Scientific Computing, Second Edition (Volume 1 of Fortran Numerical Recipes) by William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery (Cambridge University Press, 1992). This book, as well as its (optional) companion Numerical Recipes in Fortran 90: The Art of Parallel Scientific Computing, Second Edition (Volume 2 of Fortran Numerical Recipes), is available online at: www.library.cornell.edu/nr/.
Other (optional) books that students might find useful are:
- Numerical Methods in Economics by Kenneth L. Judd (MIT Press, 1998).
- Handbook of Computational Economics (Volume 1), edited by Hans M. Amman, David A. Kendrick, and John Rust (North-Holland, 1996).
- Computational Methods for the Study of Dynamic Economies, edited by Ramon Marimon and Andrew Scott (Oxford University Press, 1999).
- Dynamic Economics: Quantitative Methods and Applications by Jérôme Adda and Russell Cooper (MIT Press, 2003).
- Applied Computational Economics and Finance by Mario J. Miranda and Paul L. Fackler (MIT Press, 2002).

Grading:
The course grade will be based on two (equally-weighted) projects, one for the first part of the course and one for the second part of the course. Each project consists of writing a program in Fortran to solve an assigned problem. Students must submit their code as well as a brief (roughly five pages) description of their numerical findings. The first project will involve solving for the competitive equilibrium of a dynamic macroeconomic model; the second project will involve solving and estimating a discrete-choice dynamic programming model. Fortran is the language of choice for most researchers in computational economics; requiring that the code for the projects be written in Fortran will help students to become proficient in this powerful and useful language. The first project is due on Monday, November 14 and the second project is due at the end of the semester. Occasional short programming problems may also be assigned as the course proceeds. The purpose of these assignments is to help students develop the skills they need to complete the projects; these assignments will not be graded.

Approximate Schedule of Lectures (Part I)
I. **INTRODUCTION**
Lecture 1 Introduction to numerical dynamic programming (built around the stochasticgrowth model and the Aiyagari (1994) model). General considerations in numerical analysis: convergence, roundoff error, truncation error. Numerical differentiation.
Readings:
- Aiyagari, S.R. (1994), "Uninsured Idiosyncratic Risk and Aggregate Saving," Quarterly Journal of Economics 109, 659–684.
- Numerical Recipes: Chapters 1 and 5.7
- Judd: Chapters 1, 2, and 7.7

II. **BASIC NUMERICAL METHODS**
Lecture 2 Root-finding in one or more dimensions: bisection, secant method, Newton's method, fixed-point iteration, Gauss-Jacobi, Gauss-Seidel, Brent's method.
Readings:
- Numerical Recipes: Chapter 9

……

Note: Example of a syllabus from Yale University, in its original format. Subsections are identified using the algorithm described in this appendix.

Appendix C Texas ERC Data: Dataset Construction and Variable Definitions

C.1 Dataset Construction

We construct the ERC analysis dataset in six steps.

1. Student transcripts We start from student transcripts for all students ever enrolled in one of our seven institutions during the years for which we have syllabi data: Stephen F. Austin State University, Sam Houston State University, Texas A&M University, University of Houston-Clear Lake, University of Texas at Austin, University of Texas at Dallas, and West Texas A&M University. Student transcripts are available for the years 2011-2023 and list all courses in which a student ever enrolled, regardless of whether the course was ultimately completed. In constructing transcript-based measures of frontier exposure, we exclude special courses whose content is typically not comparable to standard classroom instruction, such as theses and independent studies. We consider students working towards undergraduate and graduate programs, in all fields.

2. Merging syllabi to transcripts We then merge syllabi to student transcripts using course codes and year. This step allows us to assign a syllabus to each transcript course and to construct student-level measures of frontier exposure. We match 51% of all courses appearing on transcripts, and we observe syllabi for 66% of the courses on the transcript of the average student. Table C1 reports these matching rates and related summary statistics.

Table C1: Matched and Unmatched Course-Years

	Matched	Non-matched
Lower division	0.155	0.056
	(0.250)	(0.165)
Upper division	0.633	0.311
	(0.393)	(0.421)
Graduate	0.212	0.633
	(0.395)	(0.469)
Enrollment	102	45
	(293)	(159)
Share matched across courses	0.508	
Avg. share of transcript matched	0.657	

Note: Characteristics of course-year pairs matched and not matched to a syllabus.

3. Merging student characteristics Next, we merge in student-level information from administrative records, including demographics and pre-college academic measures. We observe gender and race, and we merge SAT/ACT scores submitted upon application. SAT/ACT scores are available for 63% of all students in our sample. In the empirical analysis, we control for quartiles of the standardized test score distribution within our sample and for an indicator for this variable being missing. We also merge income information from FAFSA forms. We use reported family income whenever available, and reported individual income otherwise. When multiple income values are reported for the same student, we take the maximum across all reported values. In our empirical specifications, we control for indicators for zero income, missing income, income below $40K, income between $40-90K, between $90-180K, and above $180K (roughly corresponding to quartiles of the underlying distribution).

4. Merging enrollment records We next use enrollment files to recover students' academic program histories. In particular, we calculate academic program entry year using enrollment files for the years 2005-2022. Enrollment records report, for each student and year, the degree year the student is registered for (e.g., sophomore). We thus assign each student a program entry year by considering the year when they first appear in the data and the earliest observable information on degree year (for example, if a student first appears in the data in 2011 as a sophomore, we assign 2010 as the program entry year). We also use enrollment records to define students' majors. We define majors using 4-digit CIP codes and use information on a student's declared major upon enrollment, available for 98.6% of all students. For students with double majors, we treat combinations that occur with frequency above 1% as separate majors. For all other double-major combinations, we assign the first listed major. Finally, we use enrollment records to identify whether an undergraduate student ever subsequently enrolls in a graduate program in a Texas public university. In the analysis, graduate school attendance is therefore defined as enrollment in a graduate program in a Texas public university.

5. Merging graduation records We then merge graduation files to determine whether each student graduates. In the analysis, we define a student as graduating if we observe a degree completion record in the graduation files. Since graduation major information is only available for students who graduate, our baseline major definition relies on the enrollment records described above rather than on graduation records.

6. Merging earnings records Finally, we merge in wage data from Texas Workforce Commission records. These earnings data cover the years 2011-2025 and are available for 55% of the students in our data. We convert all wage figures to 2025 dollars using the Consumer Price Index (CPI). In the analysis, we consider as missing all quarters in which earnings are below $1,000. To calculate earnings over a time span after predicted graduation (for example 1-3 years after predicted graduation), we consider a predicted graduation year equal to t+6 for undergraduates and t+3 for graduate students.

Appendix D Institutional Evidence on Syllabus Visibility at Registration

This appendix evaluates a key institutional premise behind our identification strategy: whether students can observe the *updated, current-term syllabus* of a course at the time they register. The relevant empirical distinction is not whether students can ever observe any syllabus. At several institutions, students can observe either past syllabi or current syllabi after the semester begins. The identifying assumption instead requires that unexpected updates to the current-term syllabus not be fully observed by students when they choose their initial schedule. The evidence below shows that this condition is satisfied for most institutions in our sample, and at minimum is not contradicted by institutional rules for the remaining institutions.

D.1 Statewide Institutional Framework

The starting point is Texas's statewide HB 2504 regime, codified in Texas Education Code Sec. 51.974.[29] The statute requires public institutions to make undergraduate course syllabi and instructor curricula vitae publicly available online. The key point for our purposes is timing. The relevant statewide rule implementing the statute requires posting *within seven days of the first class day*, not by the opening of registration. [30] This timing rule is important because registration at Texas public universities typically opens months before the semester starts. As a result, the statewide policy itself does not guarantee that a student choosing courses during the registration window can observe the updated syllabus for the term in which she is enrolling.

The statewide rule does not imply that no instructor ever posts early. It implies the narrower and empirically relevant claim that students are generally *not guaranteed* access to the updated syllabus when they register. Several universities in our sample restate this timing explicitly on their own HB 2504 or syllabus pages. Others maintain public syllabus repositories without committing to pre-registration posting. The institution-by-institution evidence therefore varies in strength, but the modal pattern is clear: public access to current-term syllabi begins only after course registration is already underway and often only after classes begin.

[29] https://statutes.capitol.texas.gov/Docs/ED/htm/ED.51.htm#51.974.

[30] https://texreg.sos.state.tx.us/public/readtac$ext.ViewTAC?tac_view=5&ti=19&pt=1&ch=4&sch=A&rl=Y.

D.2 Interpretation for Identification

For identification, the strongest institutions are those for which university materials explicitly state that syllabi are posted *after* the first class day. At those institutions, students choosing courses during the ordinary registration period cannot observe the updated syllabus. That group includes UT Austin, West Texas A&M, Sam Houston State, Stephen F. Austin State, UT Dallas, and Texas A&M. The evidence is weaker, but still supportive in a "not guaranteed" sense, at the University of Houston-Clear Lake. At this school, we do not have institution-specific proof that the updated syllabus is always unavailable at every point in registration, but we also do not have evidence of any rule requiring public posting by the time registration opens. This is sufficient for the narrower claim that unexpected syllabus updates are not systematically observed when students initially choose classes.

D.3 Institution-by-Institution Evidence

UT Austin. UT Austin provides especially strong evidence for the identification assumption. The public syllabus and CV search page states that syllabi "become available" as they are added by faculty and departments, and the archive goes back only to Fall 2010.[31] More importantly, the registrar's registration page tells students preparing to enroll to "View past Syllabi and CVs," which is direct evidence that the institution itself does not treat updated current-term syllabi as the relevant pre-registration information set.[32] UT Austin also allows schedule adjustment after classes begin. For Fall 2025, classes began on August 25, 2025, the last day to add a class without permission was August 28, 2025, and the last day to drop without permission was September 10, 2025.[33] Historical materials show the same structure. In Fall 2012, classes began on August 29, 2012 and the official add/drop period ended on September 4, 2012; archived policy pages also allowed undergraduates to add through the twelfth class day of a long semester.[34] Taken together, the UT Austin evidence strongly supports the claim that students were not guaranteed access to the updated syllabus when

[31] https://utdirect.utexas.edu/apps/student/coursedocs/nlogon/.

[32] https://registrar.utexas.edu/schedules/262/registration.

[33] https://registrar.utexas.edu/calendars/25-26; https://registrar.utexas.edu/schedules/262; https://registrar.utexas.edu/schedules/269.

[34] https://catalog.utexas.edu/archive/2012-13/general-information/introduction/academic-calendar/; https://catalog.utexas.edu/archive/2012-13/general-information/academic-policies-and-procedures/adding-and-dropping-classes/adding-a-class/; https://catalog.utexas.edu/archive/2012-13/general-information/academic-policies-and-procedures/adding-and-dropping-classes/dropping-a-class-rules-for-undergraduate-students/.

enrolling during 2011–2022.

UT Dallas. UT Dallas is more transparent than most institutions because CourseBook is public and searchable by term back to 2002.[35] That fact, however, does not establish that the *updated current-term* syllabus must be available at registration. The current evidence does not provide such a guarantee. The Spring 2026 calendar shows that regular registration ends on January 15, 2026, classes begin on January 20, 2026, late registration and add/swap continue until January 27, 2026, and census day occurs on February 4, 2026.[36] The 2012 undergraduate catalog shows the same basic pattern: students could add without instructor or advisor signature through the sixth class day and drop without permission through census day.[37] For UT Dallas, therefore, the available evidence does not show that students were guaranteed access to the updated syllabus when registration opened or when they initially selected classes.

Texas A&M University. The current Texas A&M evidence is also consistent with the identifying assumption. The Provost's Spring 2026 start-of-semester guidance states that course syllabi must be provided to students no later than the first day of class.[38] That timing falls after most of the registration period and does not imply visibility at enrollment. The same page notes that public-facing Simple Syllabus implementation began only in August 2025 as a pilot, with broader required use later, which further suggests that historically there was no centralized requirement that students could inspect the updated syllabus before registration.[39] Texas A&M's catalog indicates that course choice can continue after the semester starts: students may enroll during the first five class days of a fall or spring semester and may drop a course with no record during the first twelve class days.[40] For Texas A&M, the evidence supports the claim that students in our sample did not have access to the updated syllabus when they chose their initial schedule.

West Texas A&M University. West Texas A&M provides direct institutional support for the claim. Its official HB 2504 page states that the university must provide undergraduate course syllabi and instructor CVs online "within a week of the first class of each semester."[41] This is precisely the

[35] https://coursebook.utdallas.edu/.
[36] https://www.utdallas.edu/academics/calendar/; https://registrar.utdallas.edu/registration/dropping-a-class/.
[37] https://catalog.utdallas.edu/2012/undergraduate/policies/registration.
[38] https://provost.tamu.edu/start-of-semester-spring-2026.html.
[39] https://provost.tamu.edu/start-of-semester-spring-2026.html.
[40] https://catalog.tamu.edu/undergraduate/general-information/grading-system/.
[41] https://www.wtamu.edu/about/information/house-bill-2504-requirements.html.

timing needed for the identification argument: students registering before classes start cannot rely on seeing the updated syllabus, because the university's own posting rule runs after the beginning of the term.[42]

Sam Houston State University. Sam Houston State offers similarly strong evidence. Its HB 2504 page posts term-specific deadlines with both the first class day and the syllabus/vita upload deadline.[43] For Spring 2026 full-term classes, for example, the page lists January 12, 2026 as the first class day and January 28, 2026 as the syllabus/vita upload deadline. This timing is well after registration and after the beginning of instruction. SHSU also made a university-wide transition to Simple Syllabus only in Fall 2025, which suggests that centralized and uniform syllabus access was, if anything, weaker during the 2011–2022 period.[44] Registration itself remains active after classes begin. For Fall 2025 full-term classes, the university's registration calendar shows multiple reopenings after the term starts, with final online registration access for the full term closing only on September 4, 2025.[45] SHSU therefore strongly supports the proposition that students could not observe the updated syllabus when they initially enrolled.

Stephen F. Austin State University. Stephen F. Austin State is another especially strong institution for the identification argument. Its Academic Affairs deadlines page states that the syllabus and CV must be uploaded for Fall 2025 courses "7 days after 1st class day," and its faculty syllabus-development page states that syllabi and current CVs must be posted by the seventh day after the first class day.[46] This directly implies that students registering before classes begin cannot observe the updated current-term syllabus as a matter of institutional policy. SFA also permits enrollment changes after classes begin. The Fall 2025 registrar calendar lists August 25, 2025 as the first class day and August 28, 2025 as the last day to register via mySFA without departmental approval; the same calendar places census later, on September 9, 2025, and indicates that late add registration begins on September 10, 2025.[47] Historical evidence points in the same direction: the 2010–11 un-

[42] https://www.wtamu.edu/academics/catalog/index.html.

[43] https://intranet.shsu.edu/dept/academic-planning-and-program-development/house-bill-2504.

[44] https://www.shsu.edu/offices-departments/shsu-online/news/stories/250820-introducing-s2-shsu-officially-launches-simple-syllabus-across-all-courses.php.

[45] https://www.shsu.edu/offices-departments/registrar/calendars/registration-open-and-close-dates; https://www.shsu.edu/offices-departments/registrar/calendars/academic-calendar.

[46] https://www.sfasu.edu/acadaffairs/calendars/deadlines; https://www.sfasu.edu/acadaffairs/curriculum/faculty-syllabus-development.

[47] https://www.sfasu.edu/docs/registrar/reg-academic-calendar-fall-2025.pdf; https://www.sfasu.edu/registrar/registration-information/how-to-register.

dergraduate catalog allowed students to add through the fourth class day in fall and spring and to drop far later in the semester.[48] SFA thus provides direct institution-specific support for our identifying assumption over the entire period of analysis.

University of Houston-Clear Lake. UHCL is the main nuance in the sample. Its HB 2504 page states that public institutions must make class syllabi and instructor CVs publicly accessible online and identifies August 31, 2010 as the effective compliance date.[49] Unlike SFA, SHSU, or WTAMU, the HB 2504 page does not itself state that the public posting deadline occurs after the first class day. The current faculty administration calendar cuts in a more complicated direction: for Spring 2026 regular-session classes, faculty submit syllabi on January 6, 2026, department chairs approve them on January 9, 2026, open registration ends on January 9, 2026, late registration and add/drop run from January 10 to January 20, 2026, and classes begin on January 12, 2026.[50] This means that students registering during much of the registration cycle likely do *not* observe the updated syllabus, but students registering very late may do so. The historical academic calendar archive is public back to Fall 2014,[51] but in the materials we collected, we did not locate a clean archived page for 2011–2013 establishing the exact timing of syllabus availability relative to registration. For UHCL, therefore, the strongest support is for the narrower claim that students were not guaranteed to observe the updated syllabus when registration opened or when they made their initial enrollment choices. Our results remain robust when we exclude students from this university from the estimation sample (Appendix Tables D1 and D2).

[48] https://www.sfasu.edu/docs/registrar/archived-catalogs/ugrad-2010-2011.pdf.
[49] https://www.uhcl.edu/policies/hb2504.
[50] https://www.uhcl.edu/provost/faculty-success/administration-calendar/.
[51] https://www.uhcl.edu/academics/resources/academic-calendar/archive/.

Table D1: Educational Effects of Frontier Knowledge Exposure - Excluding University of Houston Clear Lake

Panel (a) Undergraduate students					
	Graduates	Time-to degree	GPA	Attends grad school	Future grade rank
	(1)	(2)	(3)	(4)	(5)
Proximity (sd)	0.052***	-0.207***	0.114***	0.013***	0.009***
	(0.006)	(0.024)	(0.013)	(0.005)	(0.001)
Mean dep. var.	0.874	5.646	2.883	0.114	0.524
R^2	0.570	0.588	0.463	0.291	0.060
N	178,768	146,485	177,904	178,768	3,526,345

Panel (b) Graduate students						
	MA			PhD		
	Graduates	Time-to degree	GPA	Graduates	Time-to degree	GPA
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.014	-0.113***	0.031*	0.003	0.186	0.029
	(0.012)	(0.038)	(0.017)	(0.019)	(0.412)	(0.044)
Mean dep. var.	0.716	2.287	3.356	0.257	6.289	2.808
R^2	0.508	0.698	0.517	0.821	0.887	0.786
N	25,871	21,570	25,678	4,830	1,266	4,808

Notes: OLS estimates. Panel (a) reports results for undergraduate students and panel (b) for graduate students. The dependent variable is degree completion (column 1 in panel (a) and columns 1 and 4 in panel (b)), time-to-degree in years (column 2 in panel (a) and columns 2 and 5 in panel (b)), GPA (column 3 in panel (a) and columns 3 and 6 in panel (b)), graduate school attendance within Texas public universities (column 4 in panel (a)), and the student's rank in future course grades (column 5 in panel (a)). In all columns except column 5 of panel (a), one observation is a student; in column 5 of panel (a), one observation is a student-course pair. *Proximity* is the average frontier knowledge proximity of the courses taken by each student, measured in course-level standard deviations; in column 5 of panel (a), it is calculated using only courses taken before the focal course. All specifications control for average lagged proximity, instructor and school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Students from University of Houston Clear Lake are excluded from the sample. Robust standard errors are reported in parentheses in all columns except column 5 of panel (a), where standard errors are clustered at the student level. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.

Table D2: Earnings Effects of Frontier Knowledge Exposure - Excluding University of Houston Clear Lake

Panel (a) Undergraduate students

	ln(earnings)			Prob. earnings in...		ln(industry
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile	earnings)
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.031***	0.040***	0.026	-0.023***	0.012**	0.003
	(0.011)	(0.013)	(0.017)	(0.007)	(0.005)	(0.007)
R^2	0.297	0.302	0.353	0.215	0.314	0.276
N	178,768	148,461	80,788	178,768	178,768	178,768

Panel (b) Graduate students

	ln(earnings)			Prob. earnings in...		ln(industry
	1-6 years	1-3 years	4-6 years	bottom qtile	top qtile	earnings)
	(1)	(2)	(3)	(4)	(5)	(6)
Proximity (sd)	0.049***	0.074***	0.030	-0.020**	0.017	0.014
	(0.018)	(0.022)	(0.032)	(0.009)	(0.010)	(0.009)
R^2	0.508	0.536	0.558	0.447	0.514	0.559
N	30,428	24,964	13,556	30,428	30,428	30,428

Notes: OLS estimates; one observation is a student. The dependent variable is the natural logarithm of earnings 1-6 years (column 1), 1-3 years (column 2), and 4-6 years after predicted graduation (column 3); an indicator for earnings in the bottom quartile (column 4) and the top quartile of the graduation cohort-specific earnings distribution (column 5); and the natural logarithm of mean earnings in the first industry of employment post-graduation (column 6). Panel (a) shows estimates for undergraduate students and panel (b) shows estimates for graduate students. The variable *Proximity* is the average frontier knowledge proximity of all courses taken by each student, measured in course-level standard deviations. All specifications control for average lagged proximity, instructor and school-major-cohort fixed effects, and indicators for gender, race, family income quartile, and SAT/ACT score quartile. Students from University of Houston Clear Lake are excluded from the sample. Robust standard errors are reported in parentheses. * $p \leq 0.1$, ** $p \leq 0.05$, *** $p \leq 0.01$.