THE POLITICAL CONTENT OF COLLEGE COURSES

[CLICK HERE FOR LATEST VERSION]

JACOB LIGHT[⋆], GIDEON MOORE[‡], AND SAMUEL THAU[‡]

Abstract. We measure two dimensions of politics in college courses — the amount of political content and its partisan orientation — using course descriptions from nearly 1,000 U.S. colleges and universities over the past 25 years. Using a novel text-as-data method, we separate the amount of "political content," the extent to which a course engages with topics salient to politics and public policy, from "slant," the partisan direction of that content. Since 2000, the average course has 0.15 SD more political content and 0.13 SD more liberal slant. Changes are most pronounced in the tails of the distribution, particularly for slant, and are modest relative to persistent cross-field differences. Following instructors who move between institutions, we find that instructors account for 60-65% of cross-sectional differences in course political content and slant. Using enrollment data, we find students show no preference for liberal content before the 2010s; demand for liberal content rises over the following decade, peaks in 2020, and falls slightly thereafter.

Keywords: Higher education, curriculum, partisan slant, text embeddings

JEL Codes: I21, D72, C55

Date: August 25, 2026.

We would like to thank Peter Arcidiacono, Elliot Ash, Lanier Benkard, Camila Blanes, Nicholas Bloom, Arun Chandrasekhar, Jacob Conway, Liran Einav, Matthew Gentzkow, Paul Goldsmith-Pinkham, Jay Hamilton, Alex Hayes, Caroline Hoxby, Dan Jurafsky, Ro'ee Levy, Neale Mahoney, Raj Movva, Elena Pastorino, Ashesh Rambachan, Suproteem Sarkar, Nick Scott-Hearn, Isaac Sorkin, Janet Stefanov, Shoshana Vasserman, Sarah Vicol, Jaume Vives-i-Bastida, and Ali Yurukoglu for helpful comments. We also thank seminar participants at Stanford University and the University of Chicago. This project was supported by a grant from the Center for Revitalizing American Institutions.

[⋆] Hoover Institution, Stanford University.

[‡] Department of Economics, Stanford University.

1. Introduction

The political content of university instruction has become an increasingly prominent object of public policy. In recent years, federal and state governments have targeted university diversity initiatives, restricted funding to institutions they view as engaging in ideological advocacy, and increased oversight of university curricula. These interventions rest, at least implicitly, on empirical claims about what students encounter in college classrooms: that politics appears frequently enough to matter, that it systematically favors particular viewpoints, and that universities choose to produce it. While the policy stakes have intensified, the questions extend a debate over the existence and sources of political slant in universities that spans at least half a century (e.g., Buckley, 1951; Bloom, 1987). A recent empirical literature establishes stakes for these concerns, demonstrating the influence of higher education on political beliefs and participation (Dee, 2004; Bell et al., 2024; Firoozi, 2025; Goldstein and Kolerman, 2025; Andrews et al., 2025). Despite the prominence of these debates and evidence on the role of education in developing civic identity, researchers know remarkably little about the political characteristics of college instruction itself.

A fundamental measurement problem explains this lack of evidence. University curricula encompass hundreds or thousands of courses ranging from calculus to constitutional law, and researchers rarely observe course content at scale. Existing methods for measuring political slant primarily study settings such as congressional speech or news media (e.g., Gentzkow and Shapiro, 2010; Gentzkow et al., 2019b), where nearly every document concerns political topics. Applying such methods to college courses risks assigning partisan direction to texts that contain little political signal at all. As a result, even basic descriptive questions remain largely unanswered: How common is political content in college courses? When courses engage political topics, do they systematically lean liberal or conservative? How much have either of these margins changed over time?

We provide evidence on these questions by developing a text-as-data method that separately measures the amount of *political content*, the extent to which a course engages with topics central to politics and public policy, and *slant*, the partisan orientation of that political content. The method represents course descriptions and the 2016 Democratic and Republican Party platforms in a common semantic space and uses the platforms to identify both political relevance and relative alignment with the two partisan poles. Separating these margins allows us to classify an apolitical course as unslanted rather than forcing it onto an ideological spectrum. We validate the resulting measures using human and LLM annotations, comparisons with full syllabi, and out-of-sample political texts with known ideological positions. Importantly, slant differs conceptually from *bias*: a course may align more closely with one partisan pole because of motivated reasoning or because subject matter genuinely aligns with one party more than the other. Our measure has no concept of truth and cannot distinguish between these cases, so we adhere to positive rather than normative analysis.

We apply the method to course offerings and descriptions drawn from Light (2026). Our analytical sample includes data from 964 U.S. colleges and universities, starting as early as 2000. The data contain all observed offerings while each institution is in the sample, including section-level enrollment, instructor assignments, and course descriptions, for a total of 30.3 million course sections.

Four patterns characterize politics in college courses. First, the amount of political content and slant vary sharply across the curriculum. Social Science courses have, on average, roughly 0.8 standard deviations more political content and are 0.7 standard deviations more liberal than STEM courses, with still larger differences across individual fields within these broad groups. Second, among courses with meaningful political content, the distribution of slant skews liberal: among courses more than 1 SD away from the median slant, liberal-slanted courses outnumber conservative-slanted courses by a ratio of approximately 1.6:1. Third, courses have become modestly more political and more liberal over the past quarter century, but these changes are small relative to persistent cross-field differences: the entire 25-year shift in political content is equivalent to roughly one-quarter of the difference between the average Psychology and History course, and the shift in slant is equivalent to roughly one-fifth of the difference between the slant of the average Economics and Sociology course. Fourth, the two margins changed in different ways. Rising political content reflects growth in moderately political offerings at the expense of the apolitical left tail rather than expansion of the most political tail. The median course's slant today is very similar to the median course at the start of the sample; asymmetric movement in the tails drives the changes in mean slant.

Having established these facts, we ask what forces produce political engagement in university courses. Scholars and commentators have debated this question for decades. During the Vietnam War, The New York Times public opinion polling attributed growing politicization primarily to faculty political activism (Hitchcock, 1971); a decade later, Bloom (1987) emphasized an increasingly ideological student body to which universities had ceded their academic mission. We use two complementary empirical designs to assess contributions of the instructor and student sides of the market to equilibrium course content.

First, we focus on the influence of instructors over political content and slant. Course political characteristics represent the interplay of instructors, student demand, institutional preferences, political stakeholders, and donors, among others. By comparing course offerings of the same instructor as they move between institutions, we can isolate instructors from the collective influence of the rest of these stakeholders. Empirically, we see instructors respond proportionally when moving between more or less different institutions — justifying a movers dosage design in the tradition of Finkelstein et al. (2016). On arrival, instructors close roughly 41% of the gap in political content and 34% of the gap in slant between their origin and destination institutions. We attribute the remaining 59% of variation in political content and 66% of variation in slant to the influence of instructors.

Second, we measure how students respond to the presence of political content and slant when enrolling in courses. We estimate a nested logit model that predicts course enrollment based on political content and slant, tracking preferences for these characteristics over time.[1] Because instructors may adjust course content in response to anticipated demand, we instrument for current political characteristics using the same instructor's course content four years earlier. The lag allows student cohorts to turn over, isolating the instructor's persistent orientation from any contemporaneous demand shock. Student demand for political content is positive and stable across the sample. We find no significant preference for liberal content at the beginning of our sample. Demand for liberal courses emerges during the 2010s, peaks in 2020, and partially reverses thereafter. Thus, the changing characteristics of courses aligned with changes in underlying student demand over this period. We do not interpret this alignment as evidence that demand caused the aggregate supply trend, but it shows that changes in course offerings coincided with meaningful changes in the preferences of the students choosing among them.

The paper makes three contributions. First, and most centrally, we provide systematic evidence on an empirical object that has featured prominently in debates over higher education but that researchers have struggled to observe: politics in college instruction.[2] We characterize the prevalence

[1]Our approach builds on a literature modeling students' field and course choices as functions of human capital investment and non-pecuniary preferences (Arcidiacono, 2004; Beffy et al., 2012; Wiswall and Zafar, 2015; Ahn et al., 2024). Closest in spirit is Ahn et al. (2024), who estimate student demand for courses as a function of grading policies; we extend this approach to political characteristics and their evolution over time.

[2]Our analysis captures one specific dimension of ideological content: the political orientation of the topics a course covers, as reflected in its description. We do not measure instructor conduct, the classroom social environment, or institutional structures outside the classroom. Course descriptions are, however, the primary information available

of political content, its partisan direction, its distribution across the curriculum, and its evolution over a quarter century. Second, we introduce a text-as-data measure that separates the amount of political content from partisan slant, extending text-as-data tools to settings in which many documents contain little or no political content. Third, we provide evidence on how producer and consumer preferences jointly shape an equilibrium characteristic of higher education, connecting the paper to work on ideological content in multi-sided information markets (Gentzkow and Shapiro, 2010; Martin and Yurukoglu, 2017; Boxell and Conway, 2022; Cagé et al., 2025).

Our findings complement work that studies the political consequences of education and curricular content. Existing research shows that curricular reforms can affect political attitudes (Cantoni et al., 2017; Chen et al., 2023), while college attendance affects partisan identity and political participation (Dee, 2004; Bell et al., 2024; Firoozi, 2025), with heterogeneous effects across fields of study (Goldstein and Kolerman, 2025). Universities may also affect the politics of surrounding communities (Andrews et al., 2025). These papers largely study outcomes of educational exposure rather than directly observing the instructional content through which such effects might operate. Baum et al. (2026) provide evidence against a direct transmission channel through faculty party affiliation. Most closely related in subject, Shields et al. (2025) use Open Syllabus data to examine whether canonical texts on selected contested issues appear alongside their critics. Their hand-selected set of texts permits close study of particular debates; our approach instead measures political content and slant across the full curriculum, allowing us to characterize their distribution and magnitude systematically.

Methodologically, we build on a long literature that measures political slant in text and media content (Groseclose and Milyo, 2005; Gentzkow and Shapiro, 2010; Martin and Yurukoglu, 2017; Gentzkow et al., 2019b; Widmer et al., 2022; Jelveh et al., 2024; Braghieri et al., 2024; Cagé et al., 2025). Our approach differs from the literature by separately identifying the extensive margin of the amount of political content and the direction of partisan alignment and by using semantic

to students making enrollment and major decisions, making them directly relevant to the demand-side patterns we study.

comparisons rather than supervised classification.[3] This distinction matters particularly in settings like higher education, where most texts contain little inherently political content.

The remainder of the paper proceeds as follows. Section 2 describes the course offerings data. Section 3 introduces our measures of political content and slant. Section 4 documents their distribution and evolution over the last 25 years. Section 5 studies instructor contributions using the movers design. Section 6 estimates student demand for political characteristics. Section 7 concludes.

2. Data

We use the dataset compiled in Light (2026), which collects a panel of course catalogs and schedules from colleges and universities across the United States. The data contain course section-level detail for all courses offered for each university-year in the sample. A typical observation in the dataset includes the course number and title, section-level enrollment, instructor, and a brief text description (3-5 sentences) that summarizes the core topics covered in the course. For most of our analysis, we aggregate individual sections of the same course-term to a single observation, summing enrollment and tracking all instructors of the course.[4] We observe department names that are standardized across institutions. We refer to these standardized names as fields, as in some circumstances they may encompass multiple closely related departments. For example, the dataset assigns courses offered under separate Genetics, Microbiology, and Biological Sciences course codes to a common Biology field.

The full dataset covers course offerings from more than 1,000 colleges and universities, with observations dating back as early as 1993. The data cover institutions that enroll more than half of U.S. four- and two-year non-profit students and are broadly reflective of the U.S. higher education system. For our main analyses, we restrict attention to the 964 institutions for which we observe course descriptions. Because data coverage is more comprehensive in later years, we focus primarily

[3]Our approach draws on work using the geometry of embedding spaces to recover interpretable semantic dimensions (Bolukbasi et al., 2016; Garg et al., 2018; Ethayarajh et al., 2019; Kozlowski et al., 2019; Park et al., 2023; Huh et al., 2024). The closest measurement approaches for measuring slant are Boxell and Conway (2022) and Longuet-Marx (2024). Boxell and Conway (2022) uses text embeddings to construct one-dimensional measures of partisan slant via supervised learning. Longuet-Marx (2024) uses supervised learning to separate issues within politics, rather than measure the amount of politics at all. Magnolfi et al. (2025) and Compiani et al. (2026) both estimate demand based on product characteristics derived from embeddings, similar to our analysis in Section 6.

[4]A "section" is a specific offering of a "course" — e.g., an 8am session of Economics 101. If an institution offers two sections each of Econ 101 and Econ 102 in both the Fall and Spring semesters, the total would be 8 sections for 2 courses in Economics. Following Light (2026), we exclude non-classroom-based courses (e.g., independent study, internships).

on course offerings from the 2000-01 academic year onward. Approximately half of the institutions are observed for at least 10 years, but coverage is relatively sparse in the 1990s and early 2000s.

Subject to the restrictions described above, we observe 30.3 million course sections in our sample, with a total of 4 million unique course descriptions. We aggregate sections to yield a sample of 17.4 million unique course-by-term observations.

2.1. **Course Descriptions.** We develop a text-based measure of the amount of political content in each course and its slant based on its title and description. First, we append the title and course description into a single string. Then, we apply standard text processing to strip whitespace and standardize formatting. Finally, we strip sentences that contain "boilerplate" information, such as prerequisites or course enrollment restrictions. This processing serves two purposes: first, to distill content most relevant for our analysis of political and partisan content from the course description, and, second, to strip out differences in the way course descriptions are written at different institutions such that our comparisons focus on substance rather than style.[5]

Course descriptions capture a specific margin of course content: the information used to advertise a course in a catalog. The political characteristics we measure, therefore, reflect only what surfaces at that margin, not everything political taught in a classroom. Nevertheless, insofar as it is the information available to students making course enrollment decisions and to the extent that it correlates with what is actually taught in the classroom, we see course descriptions as an important proxy for the political content and slant of college courses. In Section 3.3, we validate the measure against matched syllabi from a single university. Because syllabi are longer and update more frequently than descriptions, their agreement with our measure indicates that descriptions capture the relevant variation despite their brevity.

2.2. **Faculty, Instructors, and Movers.** For the analysis in Section 5, we construct a panel following instructors across institutions. We produce a format-standardized list of instructors using a combination of a classification algorithm and hand-inspection, then run an iterative matching procedure to match instructors across institutions. Matches are based on name and field; we exclude any match if the instructor is teaching at multiple institutions in the same term. For details on the matching procedure, see Appendix C.2.

[5]We describe the boilerplate exclusions in greater detail in Appendix A.3.

The resulting panel includes 1.3 million unique instructors, approximately 45 thousand of whom teach at exactly two institutions over the course of our sample and have course descriptions observed on both sides of the move.

2.3. **Student Enrollment.** For the majority of the institutions in the sample, we also observe section-level enrollment and capacity. We use these values to construct enrollment shares for our analysis in Section 6. In total, we observe enrollment for 23.4 million course sections.

3. Measuring Political Characteristics of Courses

Measuring the partisan slant of text such as college course descriptions poses a challenge that does not arise in most prior work on text-based ideology measurement. Courses vary not only in the direction of their political content, but in whether they contain meaningful political content at all — therefore, a unidimensional measure of partisan slant may conflate intensity with direction. Prior approaches applied to settings like newspapers or cable news implicitly assume the texts being analyzed are uniformly political. Course descriptions, which range from calculus to constitutional law, clearly violate this assumption, which introduces substantial noise into standard approaches.

We develop a novel measure that identifies both the *extent* and *type* of political content in a course. We define the amount of *political content* as the extent to which a course engages with political or policy-relevant topics. We define *slant* as the relative alignment of this content to liberal versus conservative perspectives. Our measure compares numerical representations of the course descriptions to representations of partisan benchmarks. The reference texts define "liberal" and "conservative" content within the representation space; we measure the quantity and direction of political content within each course in relation to the benchmarks. We calculate our measures as functions of angles between course descriptions and our reference texts. In contrast, Appendix A.2.3 shows that in a stylized setting, slant measures based on supervised learning are not identified from the data alone when there are heterogeneous amounts of political content.

This section describes the construction and implementation of the measures and validates their performance against several internal and external benchmarks.

3.1. **Approach: Identifying a Political Subspace.** To separate a course's political content and slant, we develop an approach that uses semantic similarity to benchmark texts. Most numerical representations of text preserve some degree of semantic similarity. In the case of bag-of-words methods that count usage of words or phrases, shared vocabulary between texts mechanically generates similar numerical representations (Gentzkow et al., 2019a); for more complex representations like dense text embeddings, equivalence between semantic similarity of texts and closeness in embedding space is an empirical feature of most text embedding models (Bolukbasi et al., 2016; Neelakantan et al., 2022; Park et al., 2023; Huh et al., 2024; Dell, 2025).

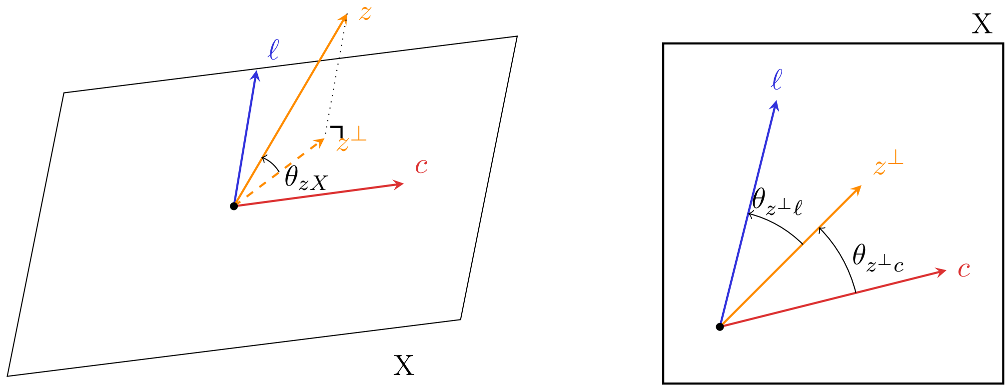
Distances between texts in representation space indicate how similar they are, but not necessarily *why* they are similar. In our setting, courses may collocate because they are political, or because they share technical or thematic vocabulary unrelated to politics. A measure built directly on proximity to political reference texts would therefore load on both political and non-political similarity, assigning apparent alignment to any course whose content happens to resemble the references for incidental reasons. To solve this problem, we isolate the portion of the representation space that captures political content and, within that space, the dimensions that distinguish partisan orientations.

The key idea is to identify the partisan regions of the representation space based on a pair of benchmark texts that are (i) unambiguously political and (ii) representative of opposing ideological poles. The span of these benchmark representations identifies the political subspace. Any course description can then be characterized by how much of its content falls within this region (amount of political content) and, among the content that does, how much it aligns with one pole versus the other (slant). Intuitively, consider comparing course descriptions to two sets of texts — one epitomizing the liberal pole of American politics, and one epitomizing the conservative pole. A course on constitutional originalism sits close to the conservative pole; a course on reproductive justice sits close to the liberal pole; a course on administrative procedure sits within the subspace but near the center; and a course on multivariate calculus sits almost entirely outside it.

To formalize the above intuition, let z denote the representation of a course description, assumed to be a unit vector in $\mathbb{R}^d$. Let $\mathbf{C}$ and $\mathbf{L}$ denote matrices whose columns are the representations of a set of conservative and liberal reference texts, respectively. Let $X = [\mathbf{C} \mid \mathbf{L}]$ denote their horizontal concatenation. Figure 1 illustrates the geometry in a simplified two-dimensional case with a single conservative reference vector c and liberal reference vector ℓ in ambient dimension three; our method generalizes to arbitrary numbers of reference vectors from each pole, as discussed in Appendix A.2.

We define the political subspace $\mathcal{X}$ as the span of X — the set of all linear combinations of the liberal and conservative reference representations. Let θ_{zX} denote the angle between z and $\mathcal{X}$. We define the amount of *political content* of a course as:

FIGURE 1. Geometric illustration of the political content and slant measures



(A) Measurement of political content of course z. (B) Measurement of slant of course z.

Notes: These figures demonstrate, in low dimensions, the geometric intuition for the political content and slant measures of a course z relative to conservative reference text c and liberal reference text ℓ. X denotes the political subspace spanned by c and ℓ, and $z^\perp$ denotes the projection of z onto that subspace. Panel (a): amount of political content is the squared cosine of the angle between z and X, equivalently the squared magnitude of the projection $z^\perp$. Panel (b): slant is the difference in squared cosines of the angles between the projected vector $z^\perp$ and each pole (c versus ℓ), scaled by amount of political content.

$$p(z \mid \mathbf{C}, \mathbf{L}) = \cos^2(\theta_{zX})$$

This measure is appealing for its geometric and statistical interpretations, as well as its robustness to representation error (Appendix A.2). Most critically, we give conditions under which measurement error in representation space becomes classical for p.

We next measure a course's *slant*. Let θ_{zC} and θ_{zL} be the angles between z and the individual subspaces spanned by $\mathbf{C}$ and $\mathbf{L}$, respectively. We define *slant* as:

$$s(z \mid \mathbf{C}, \mathbf{L}) = \cos^2(\theta_{zC}) - \cos^2(\theta_{zL})$$

To understand the mechanics of slant, let $z^\perp$ denote the projection of z onto $\mathcal{X}$, representing the component of the course representation that lies within the political subspace. Since both $\mathbf{C}$ and $\mathbf{L}$ lie within $\mathcal{X}$, the relative geometry of $z^\perp$ with respect to each pole captures ideological orientation. Define an intermediate measure, *valence*, as:

$$v(z \mid \mathbf{C}, \mathbf{L}) = \cos^2(\theta_{z^\perp C}) - \cos^2(\theta_{z^\perp L})$$

where $\theta_{z^\perp C}$ and $\theta_{z^\perp L}$ are the angles between $z^\perp$ and the conservative and liberal subspaces. Valence captures relative proximity to each partisan pole within the political subspace, but does not account for whether the course is political in the first place. A course with almost no political content could nonetheless have substantial valence if it happens that the (limited) political content is slightly more similar to one partisan pole than the other.

As shown in Appendix A.2, we can decompose slant in the following manner:

$$\begin{aligned} s(z \mid \mathbf{C}, \mathbf{L}) &= \cos^2(\theta_{zC}) - \cos^2(\theta_{zL}) \\ &= \underbrace{\cos^2(\theta_{zX})}_{p(z|\mathbf{C},\mathbf{L})} \cdot \underbrace{\left[\cos^2(\theta_{z^\perp C}) - \cos^2(\theta_{z^\perp L})\right]}_{v(z|\mathbf{C},\mathbf{L})} \end{aligned}$$

Slant is, therefore, the product of political content and valence: it captures both the magnitude of political engagement and its partisan direction. This decomposition clarifies two ways a course can have zero slant: it may be apolitical (low p), or it may engage equally with both partisan poles (low v). By convention, we define positive slant as relative proximity to the conservative pole; negative slant indicates relative proximity to the liberal pole. This scaling is important, as looking only at valence without adjusting for political content will amplify small signals in non-political courses; more specifically, courses with low political content but that engage with topics adjacent to areas of substantial ideological disagreement would introduce substantial noise if not scaled by political content.[6] Appendix A.2 gives conditions under which measurement error in s is classical.

One interpretive note: a slant value of zero, conditional on a given level of political content, should not be read as ideologically neutral. It reflects equal similarity to both poles as defined by our reference texts, not an absence of ideology. For this reason, we avoid interpreting zero as a substantive ideological center and instead focus on relative comparisons (over time, across institutions, or across fields) anchored to reference points within the data.

3.2. **Implementation.** We represent both course descriptions and platforms using OpenAI's text-embedding-3-large. Dense text embeddings are essential given the brevity of course descriptions, which can make bag-of-words representations unreliable (Neelakantan et al., 2022).[7]

[6]As an example, the course "Intro to Hispanic Linguistics" has extremely low political content, but the course has substantially liberal valence. After scaling the valence by the political content, the slant is small but liberal.

[7]Courses that cover topics "freedom from regulation" and "freedom to have an abortion" are both strongly political, but represent opposite partisan poles. A course that references "freedom of information act requests" discusses a topic that is political, insofar as it concerns government procedure, but carries no obvious partisan lean. A course in statistics that introduces the concept of "degrees of freedom" is entirely apolitical. Thus, the presence of the word "freedom" is insufficient to classify these courses. While in principle it is possible to aggregate across many words, the

Our preferred specification uses the 2016 Democratic and Republican Party platforms as the partisan reference texts. Platforms are well suited to this role for three reasons. First, platforms are deliberately comprehensive, articulating positions across the full range of political issues in an election cycle and thus spanning a broad political vocabulary. Second, using platforms avoids imposing ad hoc definitions of what counts as liberal or conservative. Third, platforms are written in a formal, programmatic style that closely resembles course descriptions — far more so than news transcripts or Congressional floor speeches, which are common benchmarks in prior work (Gentzkow and Shapiro, 2010; Martin and Yurukoglu, 2017; Gentzkow et al., 2019b; Widmer et al., 2022). Because platforms and course descriptions share this formal register, the measured political signal is less likely to be contaminated by stylistic differences between the reference and target texts. Embedding each platform section-by-section with the same model yields 109 vectors from the Democratic platform and 91 from the Republican platform.[8]

In our preferred specification, we use a fixed benchmark — the 2016 platforms — rather than contemporaneous platforms for each year.[9] This choice holds the benchmarks constant across years, so any trends we measure reflect changes in courses relative to a fixed definition of the partisan poles, rather than drift in the platforms themselves.

Given our choice of party platforms as the benchmark texts, we can further interpret the meaning of political content and slant. Political content captures the extent to which course content aligns with issues relevant in politics and policy, as specifically defined by the Democratic and Republican parties in a given time period. Our slant measure captures the ideological orientation of courses among politically salient courses through both how much each party discusses a given issue and how each party discusses a given issue. Because the 2016 platforms emphasize some topics asymmetrically, courses on those topics will inherit slant partly from the imbalance itself rather than from how they are discussed. For topics both platforms address, the measure distinguishes how each party

short length of descriptions makes the odds any given word appears small, such that these estimates become noisy to the point of inutility.

[8]To de-noise the subspace, we apply an SVD procedure. See Appendices A.2 and A.3 for an econometric treatment and discussion of our estimator. In practice, we work with dimension-reduced party platforms that are jointly 13-dimensional, and both the Republican and Democratic platforms are 9-dimensional.

[9]We have also constructed measures using platforms from 1992 onward and using platforms contemporaneous with each course year. Results are largely unchanged: political content and slant are highly correlated across benchmark choices, as shown in Figure A.6. Using a fixed benchmark has an additional advantage: temporal changes in measured course content are entirely attributable to changes in course offerings or enrollment, rather than changes in the partisan platforms themselves. Results using contemporaneous platforms are presented in Appendix A.4. We also conduct an exercise using alternative platforms as references, using the introduction to Project 2025's 2023 "Mandate for Leadership" and the 2021 Democratic Socialists of America's party platform as the conservative and liberal references respectively; results from this exercise are available in Appendix A.5.

discusses the topic: a course whose description resembles the Democratic platform's stance codes liberal, one resembling the Republican Party's stance codes conservative.[10]

3.3. **Validation.** The most direct check on our measures is whether they align with independent assessments of the same texts. Human ratings are strongly correlated with our embedding-based scores, exhibiting a Spearman correlation of 0.71 for political content and 0.48 for slant; LLM ratings are similarly strong, showing a correlation of 0.73 for political content and 0.53 for slant.[11]

Course descriptions are short and updated infrequently, raising two concerns: that their brevity may fail to capture a course's full content, and that their infrequent updates may cause them to lag changes in what a course actually covers. Full syllabi address both concerns. Syllabi are substantially longer and are not subject to the same administrative frictions that might induce stickiness. If our description-based measures agree with syllabus-based measures, neither brevity nor staleness is materially degrading the signal.

We validate our measures of political content and slant based on course descriptions against syllabi from a single university whose syllabi we can match to course descriptions. For these courses, the correlation between description-based and syllabus-based measures is high: 0.77 for political content and 0.78 for slant (Figure A.1). Critically, this is not driven by cross-field sorting: correlations are similarly high within field (Figure A.2). Moreover, the description-based and syllabus-based measures recover similar cardinal values. Since the course description is contained verbatim within the syllabus less than 0.1% of the time, this agreement reflects genuinely different language recovering the same quantity.

As an internal check on the sources of variation in our measures, we examine which phrases are most predictive of high or low political content and slant. For each two- or three-word phrase w among the ten thousand most common in the course descriptions, we compute:

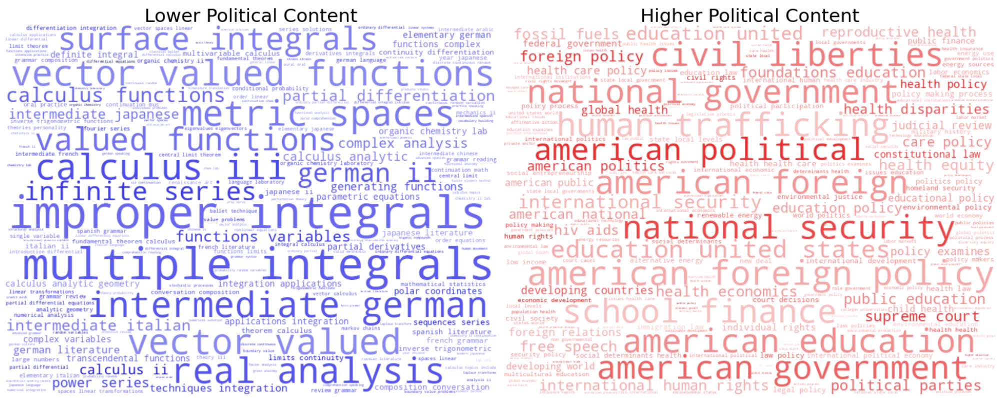
$$\mathbb{P}(\text{Statistic}(\text{Description}) \geq s \mid w \in \text{Description})$$

[10]For example, the 2016 Republican and Democratic platforms both devote substantial attention to immigration, but emphasize different aspects of the issue. The Republican platform focuses on border security, enforcement, and labor market competition, while the Democratic platform emphasizes pathways to citizenship, family reunification, and immigrant integration. As a result, a course on immigration enforcement and border security would tend to align more closely with Republican platform language, while a course on immigrant integration and citizenship would align more closely with Democratic platform language, even though both courses concern immigration.

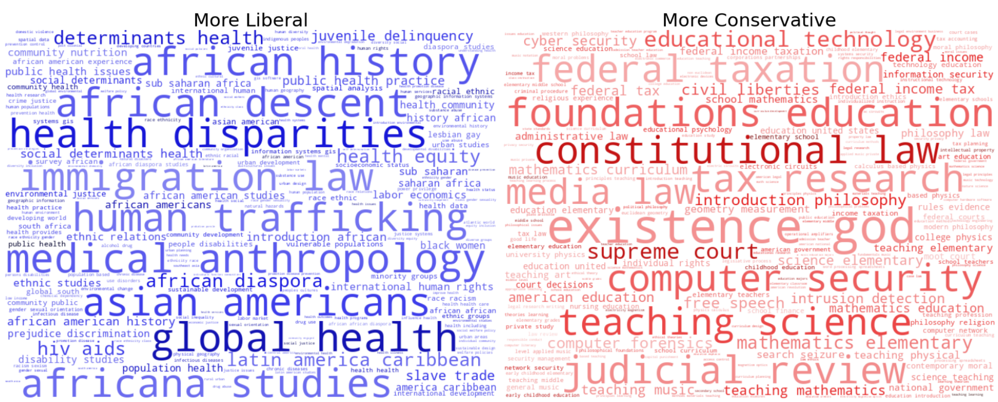
[11]For a comprehensive summary of the validation exercise, see Appendix E.

where s is the 90th percentile of either political content or slant. This allows us to capture the words associated with the "rightmost" tail (very political or very conservative). We then run the symmetric exercise looking at words associated with the "leftmost" tail (very apolitical or very liberal) by looking at frequency below the 10th percentile. As a concrete example, conditional on including the phrase "global health," we compute what fraction of descriptions have slant below the 10th percentile, corresponding to the most liberal courses. Results are shown in Figure 2.

FIGURE 2. Phrases most predictive of high and low political content and slant



(A) Phrases predicting high or low political content



(B) Phrases predicting high or low slant

Notes: Each panel is a word cloud of two- and three-word phrases drawn from the ten thousand most common phrases in course descriptions. For a phrase w, right-tail bin mass is $\mathbb{P}(\text{Statistic}(\text{Description}) \geq s \mid w \in \text{Description})$, where s is the 90th percentile of the relevant statistic for political content or conservative slant. Analogously, we compute phrases with tail bin mass below the 10th percentile of political content and conservative slant to assess the left-tail bin mass. The top panel shows phrases predicting high versus low political content; the bottom panel shows phrases predicting high vs. low (i.e., conservative vs. liberal) slant. Phrases are sized by their frequency in course descriptions and colored by their tail bin mass.

Courses with low political content are disproportionately associated with mathematics and introductory language instruction — fields with no natural connection to political themes. Courses with high amounts of political content cluster around civil liberties, international relations, and health policy. Among slanted courses, the most liberal-leaning phrases relate to health equity and the history and politics of race in the United States; the most conservative-leaning phrases relate to constitutional law and philosophical and religious concepts. These patterns confirm that the measures are sensitive to political topics rather than to incidental stylistic features of academic writing.

A phrase like "global health" is strongly associated with the liberal tail, but this does not mean the phrase is intrinsically liberal in isolation. Rather, descriptions containing it are close to the liberal platform in embedding space due to context and tone, which individual phrases cannot capture. Similarly, the conservative panel features phrases relating to K-12 education, which reflects a distinctive emphasis in the 2016 Republican platform on primary education relative to the Democratic platform, whose education statements instead emphasized higher education policy.

As a final, out-of-sample check, we apply our measures to political texts held out from the reference set: the platforms of U.S. third parties and other texts which extend ideological extremes beyond the Republican and Democratic party platforms (Figure A.3). The slant measures align with the ideological orientation of these texts — Project 2025 registers as extremely conservative and the Democratic Socialist platform registers as extremely liberal, even though they are more ideologically extreme than the partisan benchmarks.

4. The Distribution and Evolution of Political Content and Slant

Throughout this section, *political content* measures the extent to which a course engages with topics salient to politics and public policy, as represented by the 2016 Democratic and Republican Party platforms. *Slant* measures the partisan direction of that content: positive values indicate closer alignment with Republican language and negative values closer alignment with Democratic language. Because slant is scaled by political content, courses with little political content necessarily have little slant; a slant value near zero can therefore reflect either an apolitical course or a politically engaged course that aligns similarly with the two platforms.

We use these measures to characterize politics across the college curriculum. First, we document their cross-sectional distributions across fields and institutions and provide course-level benchmarks that calibrate their scale. Next, we estimate changes over the past 25 years within institution-by-field cells and examine where in the distributions those changes occur. Finally, we compare our fixed 2016 benchmark with contemporaneous party platforms to assess how changes in the parties themselves affect the measured trends.

4.1. **Static Results.** Because the measures introduced in Section 3 are novel and expressed in embedding units, we begin by documenting several descriptive patterns that help calibrate their interpretation before turning to the equilibrium analyses. Throughout, we emphasize that political content and slant capture distinct dimensions of course content. Political content measures the extent to which a course engages with political language at all, while slant measures the relative alignment of the political content with Republican versus Democratic benchmark texts. By construction, courses with near-zero political content also have near-zero slant; a slant value near zero therefore need not indicate ideological neutrality, but may instead reflect that a course contains little political content.

Table 1 provides benchmark examples that help interpret the scale of the measures. Introductory Foreign Language courses generally exhibit low political content scores, reflecting limited overlap with political language in the benchmark texts. Courses in areas such as History, Political Science, and Public Policy tend to exhibit higher political content because they engage more directly with institutions, governments, or political conflict. Conditional on being politically engaged, courses vary in slant: those focusing on topics like environmentalism and identity align more closely with language appearing in Democratic benchmark texts, while discussions of ethics and faith are more aligned with Republican benchmark texts. Most politically engaged courses remain close to the center of the slant distribution.

Table 2 summarizes descriptive statistics for political content and slant across all courses offered in the 2016-17 academic year. Recall that political content is a squared cosine value, so it is bounded between 0 and 1. The political content of the average course is 0.064. However, there exists substantial variation and the distribution is skewed by a small number of extremely political courses. Meaningful heterogeneity exists across disciplines: the average Social Science, Humanities, and Business course exhibits higher average political content, while STEM and Arts & Language courses exhibit lower political content.[12] Selective universities and those in blue states tend to have

[12]We separate Arts and Language fields (Fine Arts, Foreign Languages, and English) from the remainder of the Humanities because their offerings combine two distinct types of content with very different political profiles. Practical courses — studio technique, music lessons, language acquisition, freshman writing — are correctly identified as

Table 1. Benchmark course examples anchoring the political content and slant scales

		Political Content		Slant	
Benchmark	Course title		Norm.		Norm.
Panel A: Political Content ladder					
Political Content -1 SD	Elementary Spanish Conversation	0.0295	-1.00	-0.0021	-0.19
Political Content 0 SD	World Regional Geography	0.0640	-0.00	0.0029	+0.23
Political Content +1 SD	History, Technology, And Society	0.0988	+1.00	-0.0039	-0.34
Political Content +2 SD	American Government	0.1333	+2.00	0.0057	+0.47
Political Content +3 SD	Electoral Politics In America	0.1668	+2.97	0.0051	+0.42
Panel B: Slant ladder (among politically engaged courses)					
Slant -2 SD	Race And Ethnic Relations	0.1016	+1.08	-0.0233	-1.96
Slant -1 SD	Fund. Of Environmental Studies	0.0845	+0.59	-0.0116	-0.99
Slant 0 SD	Nonprofit Organizations & Society	0.1039	+1.15	0.0001	+0.00
Slant +1 SD	Contemporary Moral Problems	0.1037	+1.14	0.0117	+0.97
Slant +2 SD	Faith, Reason, And Culture	0.0879	+0.69	0.0242	+2.02

Notes: Each row shows the representative introductory course nearest the stated benchmark in the joint distribution of political content and slant. Panel A illustrates the political content dimension; Panel B restricts to courses 0.5 SD above the sample mean in political content (politically engaged courses) and illustrates the slant dimension. Raw values are reported in embedding units; normalized values are in standard deviation units for the full 2016-17 course sample. A slant of zero does not indicate ideological neutrality: it arises when a course is apolitical, when its content is equidistant from both party platforms, or both.

more political, more liberal courses; however, these differences are comparatively modest relative to cross-field differences.

Average slant in the full sample is essentially zero, though distributions differ across fields. STEM courses have slightly higher (more conservative) slant than the average course, while Social Sciences courses have, on average, a slant approximately 0.5 SD more liberal than the average course. Median slant for all fields is weakly higher (more conservative) than average slant, suggestive of a liberal skew. Differences in average slant across institution types and state political environments are, again, small relative to cross-field differences. Appendix Table B.2 summarizes mean and median political content and slant by field (e.g., English, Economics).

Among the minority of courses that are politically engaged, the slant distribution tends to be skewed. Figure 3 plots the joint distribution of political content and slant for four illustrative fields: Economics, Ethnic/Cultural Studies, Law, and Math.

apolitical by our measure, and vastly outnumber interpretive courses in these fields. Pooling them with History, Philosophy, and Area Studies depresses the combined Humanities average well below what readers would expect, while obscuring the concentration of liberal-skewed content in the interpretive tail. We note that the dominance of practical over interpretive courses in these fields is itself a finding.

Table 2. Descriptive statistics for political content and slant, 2016-17

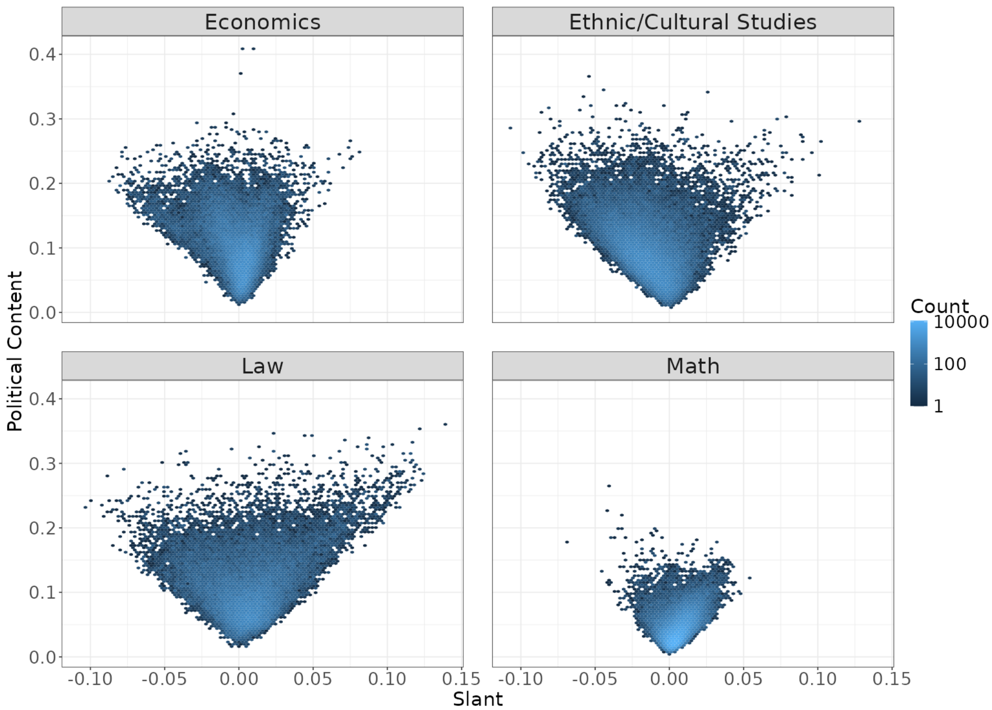
	Political Content			Slant		
	Mean	SD	Median	Mean	SD	Median
Overall	0.064	0.035	0.056	0.000	0.012	0.001
Field group						
STEM	0.054	0.027	0.048	0.002	0.009	0.002
Business	0.071	0.033	0.064	0.001	0.010	0.002
Arts & Language	0.044	0.020	0.040	0.002	0.008	0.002
Humanities	0.075	0.037	0.068	-0.001	0.016	-0.000
Social Sciences	0.082	0.042	0.074	-0.006	0.015	-0.006
Selectivity						
Highly Selective	0.068	0.040	0.057	-0.001	0.013	0.000
Mid Selective	0.064	0.034	0.055	0.000	0.012	0.001
Least Selective	0.063	0.033	0.055	0.001	0.012	0.002
Institution control						
Public	0.063	0.033	0.055	0.000	0.012	0.001
Private	0.067	0.038	0.057	0.000	0.013	0.001
State politics						
Red	0.063	0.034	0.055	0.001	0.012	0.001
Blue	0.066	0.036	0.056	-0.000	0.012	0.001

Notes: Statistics are computed at the course section level for all courses offered in the 2016-17 academic year. Political content is a squared-cosine measure bounded between 0 and 1, with higher values indicating greater engagement with political language. Slant is signed: positive values indicate closer alignment with the 2016 Republican platform and negative values closer alignment with the 2016 Democratic platform. Selectivity follows the 2015 Barron's index: Highly Selective institutions are index 1 (Most Competitive, admit rate below 15%); Mid Selective are index 2 or 3 (Highly or Very Competitive, admit rate 15-50%); all others are Least Selective. State politics classifies states as blue or red if they voted for the same party's presidential candidate in every election from 2016 to 2024; see Table B.3.

The figure illustrates substantial heterogeneity across fields in both political content and slant. Math provides the clearest example of a comparatively apolitical field, with most courses concentrated near zero political content and a correspondingly narrow slant distribution.[13] Economics occupies an intermediate position, with a moderate amount of political content and a comparatively symmetric slant distribution centered near zero. Ethnic/Cultural Studies and Law both contain substantially more politically engaged courses, though with different slant distributions: Ethnic/Cultural Studies courses are more concentrated in the negative-slant region of the support, while Law courses extend in both the positive and negative direction but have a salient right tail.

[13]The Math courses that register as relatively political are often related to mathematics education or cryptography, both of which overlap with policy-adjacent language.

FIGURE 3. Joint distribution of political content and slant for selected fields



Notes: Each panel plots the joint distribution of political content and slant for one field (Economics, Ethnic/Cultural Studies, Law, and Math). Each dot is a political content-by-slant bin, colored by the number of courses in the bin. Political content is non-negative, with higher values more political; slant is signed, with positive values conservative and negative values liberal. The triangular support is mechanical: because slant is scaled by political content, courses with near-zero political content necessarily have near-zero slant. Scales are comparable across fields.

Importantly, each field contains both relatively liberal- and conservative-coded courses, as well as courses with little political content at all.

Building on the intuition from Figure 3, Table 3 summarizes asymmetry in the slant distribution by reporting the share of courses falling more than k 2016-17 standard deviations to the right (conservative, R) and left (liberal, L) of the subgroup median, for $k \in \{0.5, 1, 2, 3\}$.

Near the center of the distribution, the asymmetry is relatively modest. At ± 0.5 SD from the median, 27.4% of courses fall in the left tail and 23.7% in the right tail. However, the asymmetry becomes more pronounced farther from the median. At ± 1 SD, the ratio of left-tail to right-tail courses is roughly 1.62:1; by ± 3 SD, the ratio has risen to 2.75:1. These patterns suggest that differences in aggregate slant are driven less by the typical course than by asymmetries among the relatively small set of courses with more extreme slant values.

TABLE 3. Asymmetry in the slant distribution: tail shares by subgroup

		±0.5 SD		±1 SD		±2 SD		±3 SD	
	Median	L	R	L	R	L	R	L	R
Overall	0.001	27.4%	23.7%	14.8%	9.1%	4.1%	1.5%	1.1%	0.4%
Field group									
STEM	0.002	23.1%	21.0%	10.2%	6.7%	1.7%	0.6%	0.2%	0.1%
Business	0.002	23.6%	20.5%	11.2%	6.1%	2.3%	0.8%	0.7%	0.1%
Arts & Language	0.002	20.4%	18.4%	8.0%	4.4%	1.4%	0.3%	0.2%	0.0%
Humanities	-0.000	34.0%	32.7%	22.2%	18.4%	8.3%	4.2%	2.4%	0.8%
Social Sciences	-0.006	32.3%	31.6%	18.6%	16.0%	5.4%	3.9%	1.5%	1.4%
Selectivity									
Highly Selective	0.000	28.7%	24.0%	16.3%	9.1%	5.3%	1.7%	1.7%	0.5%
Mid Selective	0.001	27.5%	23.8%	14.9%	9.2%	4.0%	1.6%	1.0%	0.4%
Least Selective	0.002	26.8%	23.4%	14.1%	8.8%	3.7%	1.5%	0.9%	0.4%
Institution control									
Public	0.001	27.3%	23.5%	14.6%	8.8%	3.9%	1.4%	1.0%	0.3%
Private	0.001	27.6%	24.2%	15.3%	9.7%	4.6%	1.9%	1.4%	0.5%
State politics									
Red	0.001	26.9%	23.5%	14.2%	9.0%	3.6%	1.5%	0.9%	0.4%
Blue	0.001	28.0%	24.1%	15.6%	9.3%	4.6%	1.6%	1.3%	0.4%

Notes: Each cell reports the percent of courses with slant more than k 2016-17 standard deviations to the right (R, conservative) or left (L, liberal) of the subgroup median, for $k \in \{0.5, 1, 2, 3\}$. Tail thresholds use the pooled 2016-17 standard deviation of slant ($\sigma = 0.012$) rather than subgroup-specific standard deviations, so differences across rows reflect differences in the distributions themselves rather than within-group rescaling. The median column reports the subgroup median in raw embedding units. A course is classified as conservative (liberal) if its slant exceeds (falls below) the subgroup median by more than $k\sigma$. The sample covers all course sections offered in 2016-17. Selectivity follows the 2015 Barron's index: Highly Selective institutions are index 1 (Most Competitive, admit rate below 15%); Mid Selective are index 2 or 3 (Highly or Very Competitive, admit rate 15-50%); all others are Least Selective. State politics classifies states as blue or red if they voted for the same party's presidential candidate in every election from 2016 to 2024; see Table B.3.

Tail mass differs across subgroups. Social Science fields exhibit comparatively large tail mass, while STEM, Business, and Arts & Language exhibit relatively little mass in either tail. For example, roughly 0.3% of STEM courses lie more than ±3 SD from the subgroup median, while 2.9% of Social Science courses do. The relative sizes of the left and right tails also vary across fields, though in all cases there is more mass in the liberal tail than the corresponding conservative tail. These comparisons reinforce the broader pattern from Table 2: cross-field differences in both political dimensions are substantially larger than differences across institution types.

Taken together, these patterns establish four facts that motivate the remainder of the paper. First, political content and slant are concentrated in a minority of courses and highly uneven across fields. In most courses, political content and slant are negligible. Second, among politically engaged courses, the distribution of slant is left-skewed in the aggregate. Third, the degree and character of that skew varies substantially across fields, with the Social Sciences driving most of the aggregate imbalance. Fourth, because political content and slant are so unevenly distributed across fields, differences in field composition are a natural candidate to explain cross-institutional variation in average slant — institutions weighted toward high political content fields will register as more liberal regardless of within-field orientation. We return to this decomposition in Section 5.

4.2. **Trends in Political Content and Slant.** The static patterns documented above establish that political content is modest on average and concentrated in a minority of fields and courses, while slant among the most politically engaged courses skews liberal. A central claim in the contemporary policy debate, however, is not about levels but about trends: that universities have become increasingly political and increasingly liberal over time. We examine this claim by estimating within-institution-by-field changes in political content and slant over the past two and a half decades. The within-institution-by-field design absorbs all time-invariant differences across institutions and fields — including the large compositional differences documented above — so that the estimated trends reflect changes occurring within a given subject at a given institution rather than shifts in the mix of courses across fields or institutions.

Formally, we estimate:

$$y_{cift} = \kappa_{if} + \sum_{s=2000,\, s\neq 2016}^{2025} \beta_s \, \mathbb{1}\{t = s\} + \varepsilon_{cift}$$

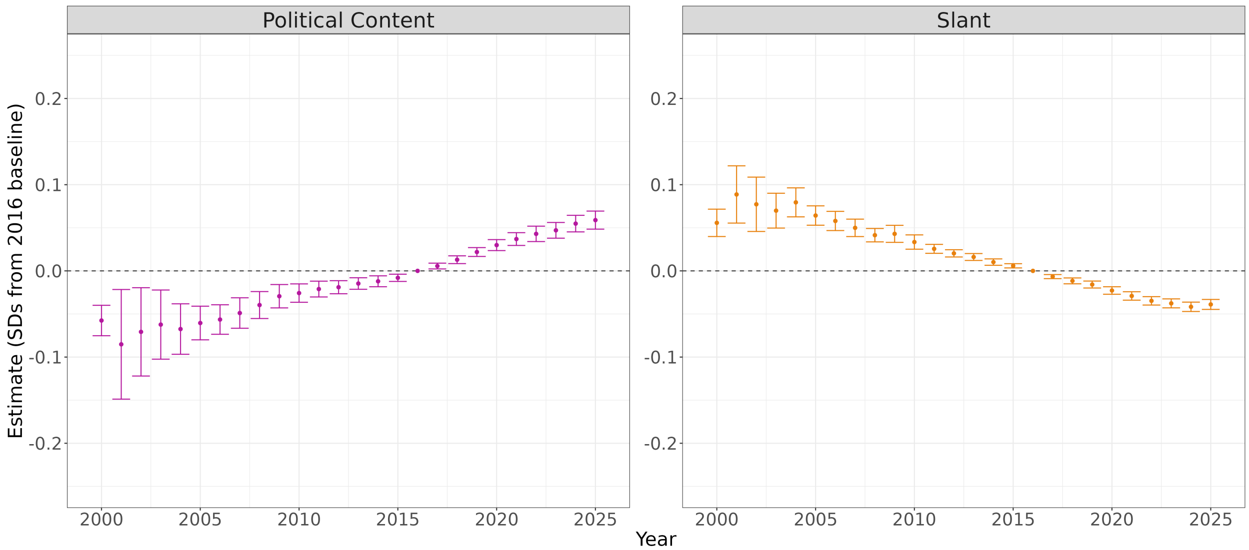
where y_{cift} is political content or slant for course c at institution i in field f in year t, κ_{if} is an institution-by-field fixed effect, and the coefficients β_s capture deviations from the 2016 baseline within each institution-field cell. Standard errors are clustered at the institution level.

To facilitate interpretation, we report estimates in all figures and tables hereafter in standard deviation units based on the distribution of courses offered in 2016-17.

Figure 4 plots the estimated $\hat{\beta}_s$ coefficients for political content and slant. Within institution-field cells, courses have become modestly more political and more liberal over the past 25 years. Political content increases approximately linearly across the sample period. The slant trend is similarly stable, with the average course trending more liberal over time.[14] The magnitudes, however, are

[14] Course descriptions update infrequently, raising the concern that the results reflect infrequent description updating rather than content stability. Our syllabus validation exercise in Section 3.3 demonstrates strong within-course correlation between descriptions and syllabi, which update more frequently, indicating descriptions track content changes relevant to our measures.

FIGURE 4. Trends in political content and slant



Notes: Points are coefficients from a regression of political content (left panel) and slant (right panel) on year indicators, controlling for institution-by-field fixed effects. Courses offered in 2016-17 are the omitted baseline, so each coefficient is a deviation from the 2016 within-institution-by-field mean, measured in 2016-17 course standard deviation units. Political content measures the extent to which a course engages political language; slant measures the partisan orientation of that engagement, with negative values indicating liberal alignment. Standard errors are clustered at the institution level; vertical bars are 95% confidence intervals.

modest. In 2016-17 cross-course standard deviation units, the 25-year change in political content amounts to a 0.15-standard-deviation increase and the 25-year shift in slant amounts to a 0.13-standard-deviation shift in the liberal direction. For reference, a 0.15-standard-deviation increase in political content is roughly one-quarter of the distance between the average Psychology and History course (Table B.2). Similarly, a 0.13-standard-deviation shift left is roughly one-fifth of the difference between the slant of the average Economics and Sociology course.

In Appendix B.2, we decompose the change over time into within-course changes, entry, and exit. The increase in political content is primarily driven by courses with a moderate amount of political content exiting and the entry of courses with even higher political content. The shift in the liberal direction is driven by left-leaning courses being replaced by more liberal courses. For both characteristics, the change within courses is small.

Table 4 summarizes these trends quantitatively, reporting fitted values in 2000-01 and 2025-26 alongside the estimated 25-year rate of change, for each subgroup. We estimate a linear version of the preceding regression, still controlling for institution-by-field fixed effects. Three patterns emerge from the table. First, the increases in political content and liberal slant are broad-based. Every subgroup exhibits a positive political content trend and a trend toward more liberal slant over the sample period. Second, the magnitude of the 25-year trend is small relative to the cross-sectional differences documented in Table 2: the persistent gap in mean political content between Social Sciences and STEM in any given year ($\approx$ 0.8 SD) is more than five times larger than the total within-field increase in political content over 25 years ($\approx$ 0.15). This comparison holds within fields as well — even in Humanities, where the trend is largest, the 25-year change is small relative to baseline cross-field differences.

Third, the trends are not uniform. Political content has increased by similar amounts across most fields other than Business. Social Sciences and Humanities exhibit the largest increases in liberal slant; STEM, Arts & Language, and Business exhibit smaller changes in slant. Among institution types, selective universities exhibit larger increases than less selective peers.

It is possible these trends in average course characteristics reflect an expanding choice set: the same conservative or apolitical courses still exist, but there are now additional political and liberal options. If this were the case, we would expect an increase in the variance of course offerings over this same period. Figure B.1 documents that there is an increase in the variance of political content and slant since 2017 within field-by-institution bins. However, we find no change in variance of valence, which suggests trends in both political content and slant are reflective of more variance in the amount, rather than direction, of politics.

4.3. **Distributional Changes.** The previous section documents growing political content (0.15 SD) and liberal slant (0.13 SD) of the average course between 2000 and 2025. This section asks where in the distribution these shifts originate: whether course offerings as a whole have become uniformly more political and liberal, or whether the trend is concentrated in specific parts of the distribution. To characterize how the distribution moves, we pool observations from 2005-10 and 2020-25 into early and late windows, respectively, and estimate movements throughout the distribution over time.[15]

Figure 5 plots the distributions of our measures, scaled to 2016 standard deviations, overlaying the 2005-2010 distribution in green and the 2020-2025 distribution in blue. The top panel of each

[15]Appendix Table B.5 tests whether the distributional shifts are statistically significant by, for each quantile cutoff in the early distribution, estimating the fraction of courses in the late distribution below it. All shifts we describe in this section are statistically significant at the $p < 0.01$ level. 2005-2010 is preferred to 2000-2005 for this exercise because the nonparametric analysis requires more data.

TABLE 4. Fitted levels and 25-year trends in political content and slant, by subgroup

	Political Content			Slant		
	Means		Δ (25 yr)	Means		Δ (25 yr)
	2000-01	2025-26	Slope	2000-01	2025-26	Slope
Overall	—	0.15	0.15	—	-0.13	-0.13
Field group						
STEM*	—	0.18	0.18	—	-0.07	-0.07
Business	0.59	0.64	0.05	-0.00	-0.11	-0.11
Arts & Language	-0.27	-0.09	0.18	0.01	-0.10	-0.11
Humanities	0.59	0.82	0.22	-0.13	-0.43	-0.30
Social Sciences	0.87	0.99	0.12	-0.62	-0.78	-0.16
Selectivity						
Highly Selective*	—	0.28	0.28	—	-0.21	-0.21
Mid Selective	-0.04	0.11	0.15	0.05	-0.09	-0.14
Least Selective	-0.03	0.07	0.10	0.09	-0.01	-0.09
Institution control						
Public*	—	0.13	0.13	—	-0.11	-0.11
Private	0.07	0.27	0.21	0.03	-0.15	-0.18
State politics						
Red*	—	0.15	0.15	—	-0.11	-0.11
Blue	0.06	0.22	0.15	-0.05	-0.21	-0.16

Notes: Fitted values in 2000-01 and 2025-26 and the estimated 25-year rate of change come from separate regressions of political content and slant on a linear year term, controlling for institution-by-field fixed effects and estimated separately for each subgroup. Fitted values are predictions from the regression evaluated in 2000-01 and 2025-26; the 25-year trend is the estimated linear slope multiplied by 25. All values are in 2016-17 standard-deviation units. Within each panel, changes are reported relative to the level of the first category (indicated with an *) in 2000-01. Standard errors, clustered at the institution level, are suppressed; all slope estimates are significant at the 1% level. Table B.4 reproduces this table with standard errors.

subfigure shows overlapping density plots, where the colored sections indicate differences between the 2005-10 distribution and the 2020-25 distribution. Each middle panel locates reference courses from Table 1 to benchmark the standard deviation units. Each bottom panel computes the net change in density at each point in the distribution between the early and late periods.

Relative to the beginning of our sample, recent course offerings contain fewer apolitical courses, replacing them with moderately political courses (Figure 5a). Most losses occur around courses with political content comparable to “Elementary Spanish Conversation.” Growth in the medium political content part of the distribution — between “World Regional Geography” and “American Government” — offsets these losses. As a result, the share of courses below the early-period median falls to 45.8%. The upper tail, by contrast, is essentially unchanged. The rise is a compression of the bottom rather than an extension of the top.

Over our sample period, mass from the center and right tail of the slant distribution has redistributed to the left tail (Figure 5b). 4.8% of the mass to the right of the early-period median has moved to the left of the early-period median. While this has only moderate effects in the thick parts of the distribution, it creates significant changes in the tails: there has been a 26.2% increase in mass in the bottom (most liberal) ventile of slant — roughly, those to the left of "Race and Ethnic Relations." There has been a smaller but still significant 11.9% reduction in mass in the top (most conservative) ventile of slant — roughly, those to the right of "Faith, Reason, and Culture." Notably, much of the mass loss from the early period comes from courses near the center of the slant distribution — those similar to "Nonprofit Organizations & Society."[16]

Together, these patterns show that the aggregate trends understate how little the typical course has changed. Political content rose across the board from the bottom up: the least political courses became scarcer while the most political courses did not multiply, so the floor rose without a new ceiling. The average course has become more liberal not just due to a fall in the availability of conservative offerings, but also due to a decline in unslanted content.

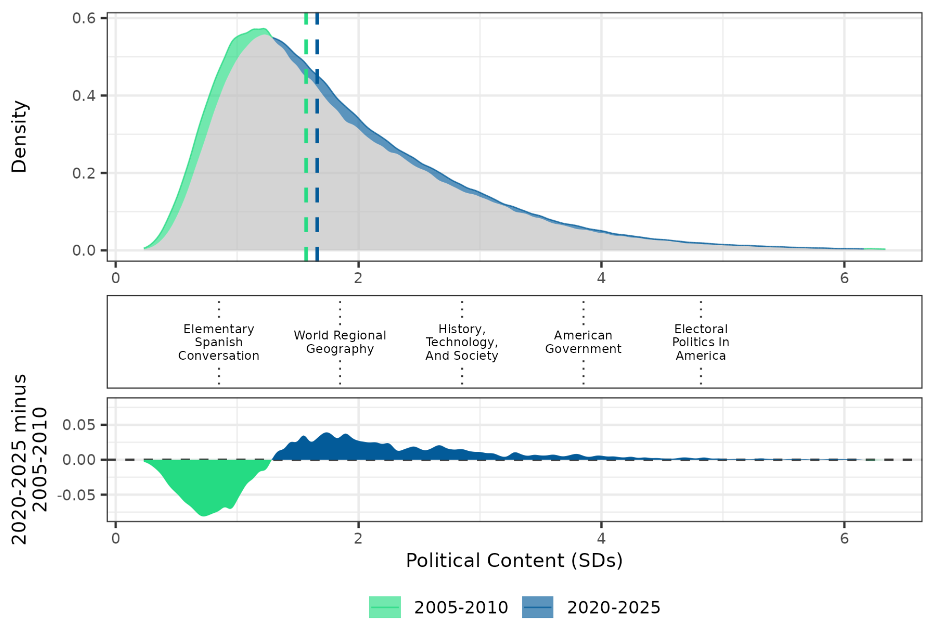
4.4. **Partisan Realignment.** Our preferred specification holds politics fixed relative to the 2016 platforms. Measuring political content and slant relative to a fixed benchmark isolates movements that come from changes in course offerings: a professor who started teaching their trade course in 2000 likely did not design their class anticipating partisan splits on trade policy in 2025. The 2016 platforms are a natural anchor, given that the period of public scrutiny of higher education that motivates this paper originates during the 2016 election. In this section, we consider an alternative specification in which politics is allowed to vary alongside courses using the platform *closest* to each course rather than a constant platform. The resulting trend therefore reflects both movement in course content and movement in the platforms themselves.

We assign each course to the platform from the nearest election year. Because different periods of time are now measured against different reference platforms, we allow institution-by-field fixed effects to vary by platform ($\kappa_{ifp(t)}$), so that demeaning occurs within each platform. For courses in

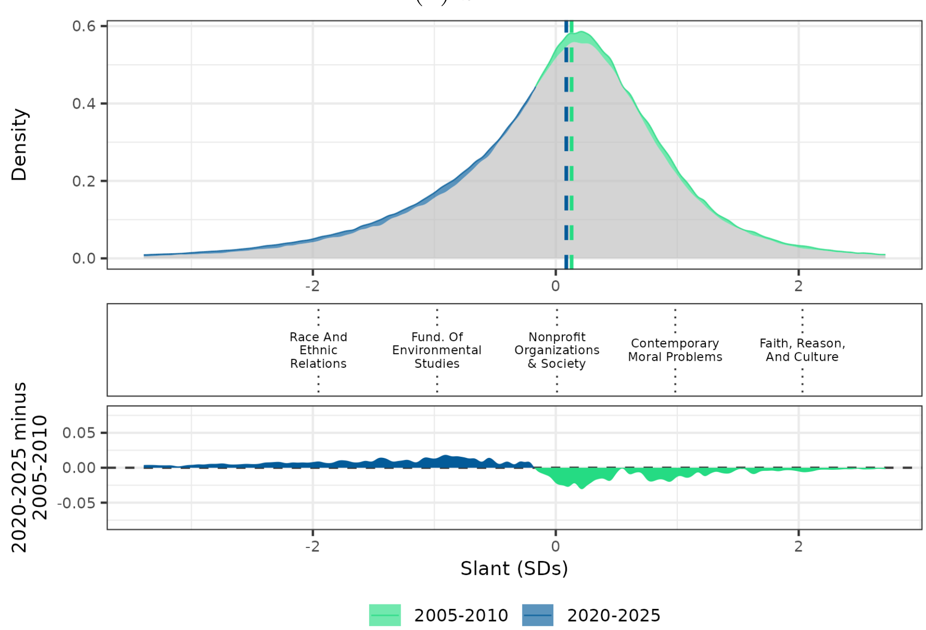
[16]Sample coverage expands over this period from 192 institutions to 962 institutions, so the raw comparisons reflect both within-institution change and the changing composition of institutions and fields. These comparisons differ from the estimates we show in Section 4.2, which control for institution-field fixed effects. For robustness, we show the same distributional shift for the balanced panel of 192 institutions in Figure B.3; the results are qualitatively similar.

Figure 5. Distribution of political content and slant for courses 2005-10 vs. 2020-25

(a) Political content



(b) Slant



Notes: Each panel compares the distribution of a course content measure — political content (panel a) and slant (panel b) — between the early (2005-2010, green) and late (2020-2025, blue) windows, in 2016-17 cross-sectional standard-deviation units. For slant, positive values are conservative and negative values liberal. Within each panel, the top sub-panel overlays the two kernel density estimates; the dashed vertical lines mark the period-specific medians. The middle sub-panel locates representative courses from Table 1 along the same axis for reference. The bottom sub-panel plots the late-minus-early density difference; positive (blue) regions gained mass and negative (green) regions lost mass. The unit of observation is a course, collapsing contemporaneous sections to a single observation.

boundary years between platforms (e.g., 2018 for the transition from the 2016 to 2020 platforms), we include two observations: one based on comparisons to the previous platform and one based on the new platform, which allows us to identify the fixed effect terms.

Recall that our slant measure is cardinal. A course with slant 0 does not mean a course is ideologically neutral, so the more appropriate comparison is relative to other courses rather than in absolute terms. Because the location of 0 is an artifact of the reference platforms we choose, in our rolling platform specification, the "location of 0" changes. Therefore, we de-mean within each platform using the fixed effects $\kappa_{ifp(t)}$. We standardize within each platform period using the standard deviations of course political content and slant in the year of the platform (e.g., 2024-25 courses for the 2024 platform).

These modifications yield the following regression specification to estimate trends in course political content and slant:

$$y_{cift} = \kappa_{ifp(t)} + \sum_{s=2000, s\neq 2016}^{2025} \beta_s \mathbb{1}\{t = s\} + \varepsilon_{cift}$$

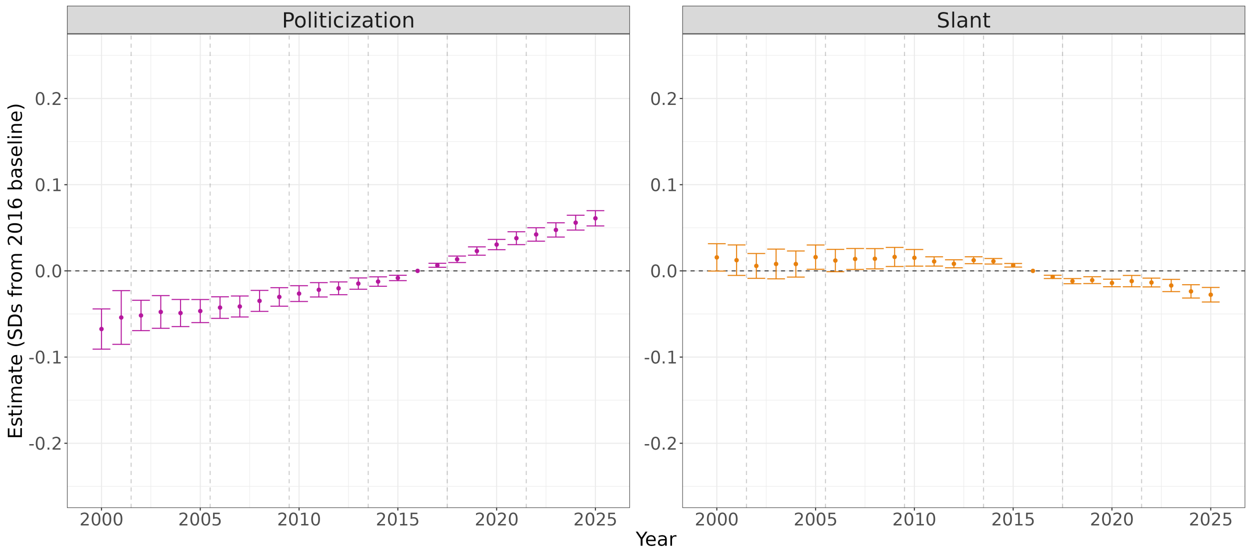
We plot regression estimates in Figure 6.

Comparing Figure 6 to the static platform specification in Figure 4, the political content trend is roughly preserved, both in shape and magnitude. The slant trend, by contrast, is substantially attenuated under the rolling specification: instead of a gradual liberal drift, slant is nearly constant through the mid-2010s, then declines sharply.

The rolling specification confounds two sources of cross-year variation that we cannot separate. First, the platforms differ across years in the topics they cover, reflecting differences in the issues that most defined the parties at that time. The 2008 platforms, for example, emphasize the economic and foreign policy issues most salient in that election, while the 2016 and later platforms center cultural issues that have greater overlap with the topics covered in college courses. Second, the parties evolve in the position they stake out on a fixed set of topics — realigning on trade, for example, without changing how much space they devote to it.

To understand the divergence in the slant trends, it is instructive to reintroduce valence: the course's relative proximity to the Republican versus Democratic platform, not scaled by political content. Figure B.2 reproduces Figure 4, including valence; the trends in valence and slant are extremely similar. However, this is not the case in the rolling specification. Figure A.7 shows

Figure 6. Trends in political content and slant, rolling (contemporaneous) platforms



Notes: Points are coefficients from a regression of political content (left panel) and slant (right panel) on year indicators. Slant and political content measures are constructed relative to the party platform closest to the year the course was offered. Analogous to Figure 4, we control for institution-by-field-by-comparison-platform fixed effects. During years when the platform switches, we include one observation based on comparisons to the previous platform and one based on the new platform, which allows us to identify the fixed effect terms. Standard errors are clustered at the institution level; vertical bars are 95% confidence intervals.

that the valence trend under the rolling specification looks similar to the slant trend in the fixed benchmark specification.

The slant divergence operates through the covariance between political content and valence: the degree to which the more politically engaged courses are also the courses that read as liberal. Under the 2016 benchmark, the covariance between political content and valence falls; that is, courses are becoming more political precisely along the cultural and identity dimension on which they also register as liberal, so political content and valence reinforce each other and mean slant drifts left. Under rolling platforms, earlier courses are measured against platforms spanning a different set of issues, where political content and liberal valence are not aligned in the same way; the covariance is less negative and the slant trend attenuates. The leftward slant trend in our preferred specification therefore reflects movement of courses along the specific partisan dimension the 2016 and subsequent platforms span — the dimension where the contemporary partisan debate plays out. A benchmark anchored to a different dimension would register the same underlying course changes differently, or not at all.

We retain the 2016 platforms as our primary benchmark for two reasons. First, the 2016 platforms emerged from an election cycle that crystallized education-based partisan realignment and elevated higher education into political contention; they capture the partisan content along which subsequent debate has played out (e.g., Zingher, 2022). Second, the 2016 platforms span a topical range that aligns substantively with the content of contemporary college courses, allowing the measure to register movements that earlier platforms — focused on different policy emphases — would not. For the remainder of the paper, we return to the 2016 specification.

5. Isolating Instructors Using Movers

Section 4 shows that political content and slant vary substantially across courses. A combination of instructor ideology, student demand, institutional culture and policies, and broader external forces contribute to these characteristics, to varying degrees, in equilibrium. Disentangling these forces is difficult in observational data where instructor, student, and institutional characteristics are jointly determined. We exploit instructor moves across institutions to isolate the role of faculty.

Consider an instructor who moves between institutions. Their own preferences travel with them, but their environment — students, colleagues, administrators — does not. The extent to which their courses change reveals the relative influence of instructors and their institutional environments. If the instructor teaches identical courses before and after the move, their environment has little influence over what they teach; if, instead, their courses shift fully to the norms of their destination, then the environment dominates. The extent of adjustment, therefore, identifies the share of course content attributable to institutional environments rather than to instructors themselves.

5.1. **Empirical Design.** We estimate the share of each political characteristic attributable to instructors relative to their institutions by formalizing the exercise above. Let y_{jt} be the average political characteristic (either political content or slant) of instructor j's courses in time period t. The instructor moves in year $m(j)$.

We call the difference between the origin and destination programs (institution-fields) the move distance, denoted δ_j. Formally, δ_j is defined as the leave-own-courses-out difference in the average value of the characteristic between the instructor's destination and origin institutions, within j's field, at the time of the move. For example, if an Economics instructor j moves from the University of Chicago to Stanford in 2017, their slant dosage is the difference in average slant of Economics courses at Stanford in 2017 relative to Economics courses at the University of Chicago in 2016, excluding their own courses. The leave-out construction ensures the dosage is not mechanically related to the instructor's own content. We assume that institutional impacts on instructors in response to the move are proportional to δ_j at rate α:

$$y_{j,t=m(j)} - y_{j,t=m(j)-1} = \alpha\delta_j$$

By dividing the change in course characteristics by the distance of the move, we can recover α: the relative influence of non-instructor factors at an institution on the characteristics of its courses. Assuming that $\alpha \in [0,1]$, the complement of this quantity is the influence of the instructor on course characteristics.

If instructors do not change their courses after moving, $\alpha = 0$, corresponding to instructors having complete control. At the other extreme, full homogenization to their destination implies $\alpha = 1$. We validate the proportional influence assumption by testing for a relationship between move distance δ_j and "on-impact" change in course content. This exercise directly estimates the average environmental influence α. Figure C.1 confirms the relationship between move distance δ_j and the change in course characteristics is approximately linear, suggesting our proportional influence assumption is well founded.

In Appendix C.3, we propose a microfoundation of the proportional pass-through assumption via a Nash Bargaining model. In the model, instructors, students, administrators, and other stakeholders in the school each have an ideal level of y. In equilibrium $y_{j,t=m(j)}$ is a convex combination of instructor ideal points and a composite ideal point representing the institutional environment. In this model, α represents the total bargaining weight on the non-moving-instructor factors, while $1 - \alpha$ is the instructor's bargaining weight.

To estimate the average α, we specify a dosage event study movers model, adapting the framework from Finkelstein et al. (2016):

$$y_{jt} = \nu + \gamma_j + \eta_{f(j)} + \tau_t + \beta \text{ length}_{jt} + \sigma_{t-m(j)} + \alpha_{t-m(j)}\delta_j + \varepsilon_{jt}$$

where y_{jt} is the average of the characteristic of instructor j's courses in term t and γ_j, $\eta_{f(j)}$, and τ_t are instructor, field, and academic year fixed effects, respectively. The $\sigma_{t-m(j)}$ terms are event-time-relative-to-move fixed effects, which absorb changes in teaching mechanically associated with career transitions, such as shifts in course load or teaching responsibilities around tenure. We control for description length length_{jt} to account for variation in style across institutions.[17] The coefficients of interest, $\alpha_{t-m(j)}$, trace out how instructors' course content evolves relative to move dosage. Changes at the time of the move identify the institution's influence on instructors. The post-period coefficients allow for dynamic effects — instructors' content may stabilize with a lag, or instructors may revert toward their own baseline over time.

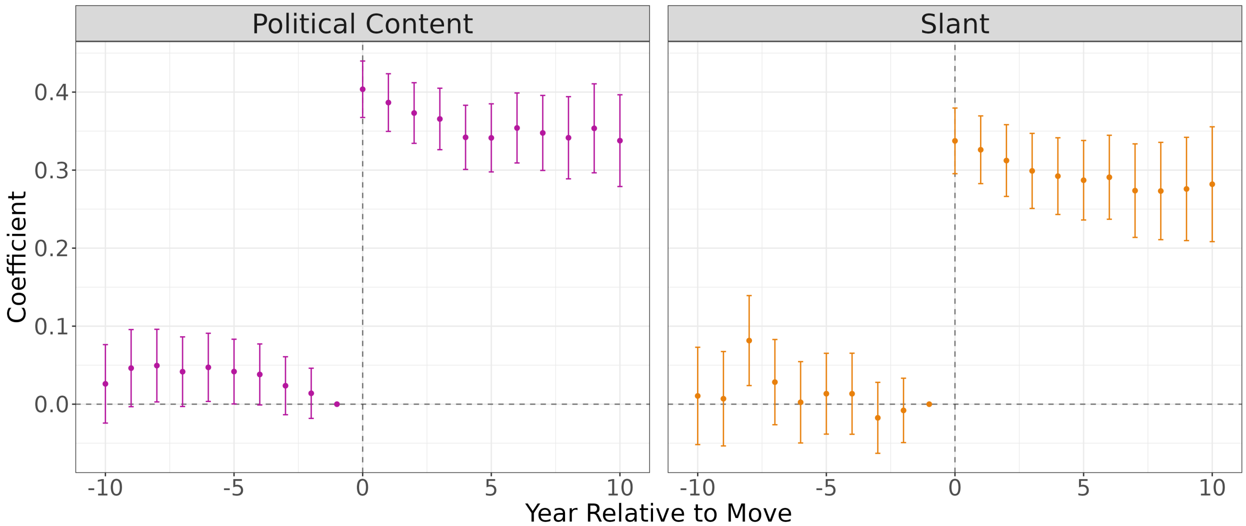
Identification relies on two assumptions: instructors do not sort systematically across institutions based on *time-varying* preferences for either political characteristic (i.e., moves are not anticipated by pre-move drift toward the destination), and instructor preferences are otherwise stable around the move, so that post-move changes reflect changing context rather than changes in the instructors themselves. Assortative matching between instructors and institutions based on stable political preferences does not threaten identification. γ_j absorbs any time-invariant component of this sorting. Identification requires only that the *timing* of moves is uncorrelated with time-varying changes in instructor preferences for political characteristics. Pre-move trends in $\alpha_{t-m(j)}$ diagnose the first assumption; $\sigma_{t-m(j)}$ addresses the second assumption by absorbing changes in teaching that may coincide systematically with career transitions, isolating the differential response to destination versus origin content.

The estimates from this exercise characterize relative influence on residual cross-sectional variation in course political characteristics, adopting the interpretation of the dosage mover design from Finkelstein et al. (2016). The exercise is not a variance decomposition – instead, it identifies the contribution of institutional environments and instructors to differences in levels, conditional on field and other fixed effects.

5.2. **Results.** Figure 7 plots the estimated influence of institutional environments $\alpha_{t-m(j)}$ for political content and slant. Pre-move coefficients are small and pre-trends are flat, showing no evidence of anticipatory drift toward the destination institution. At the time of the move, we observe discrete changes in both political content and slant in the direction of the destination institution. We estimate $\hat{\alpha}_0 = 0.41$ (s.e. 0.02) for political content and $\hat{\alpha}_0 = 0.34$ (s.e. 0.02) for slant, with changes persisting for years after the move. These estimates attribute approximately 35-40% of residual cross-sectional variation in political characteristics to non-instructor factors, leaving 60-65% attributable to instructor-specific factors.

[17] Results omitting this control are substantively unchanged.

FIGURE 7. Event study estimates of instructor adaptation around moves



Notes: Points are coefficients on event time relative to move indicators interacted with move dosage (the leave-own-courses-out difference in average political characteristics between the instructor's destination and origin institutions, within field, at the time of the move), estimated separately for political content and slant. A positive value indicates that instructors' courses become more similar to their destination institution after a move, with a value of 1 indicating full assimilation of institutional characteristics and 0 indicating no environmental impact. Standard errors are clustered at the instructor level.

Appendix C includes tests of two robustness concerns. First, since instructors who move may inherit existing courses (and course descriptions) at their destinations, content could shift mechanically toward the destination without any genuine input from the instructor, overstating the role of context. Restricting the sample only to course descriptions being offered for the first time, as in Figure C.2, decreases the $\hat{\alpha}_0$ estimates modestly (around 0.33 for political content and 0.25 for slant). This implies instructors have significantly more influence when setting their own course content rather than inheriting an existing course. We prefer the main specification, since inducing instructors to teach existing courses is itself a channel through which institutions can exert their influence. Second, to reduce the risk of false positive matches, we restrict to instructors who we observe in each of the three years before and after the move (Figure C.3)[18]; results are substantively unchanged.

[18]Imbalance in the institutional panel and incomplete coverage of all institutions may generate a gap between when an instructor leaves their origin (which we observe exactly as the year the instructor disappears from the instructional roster) and when they appear at their destination.

Appendix Table C.1 shows heterogeneity in the mover analysis by field group, selectivity, time-of-move, institution control, and geography.[19] Environmental influence is largest at the least selective institutions; equivalently, instructors at more selective institutions exert more control over their courses. These differences attenuate when using only descriptions seen for the first time (Table C.3), reaffirming course inheritance as a channel for institutional influence.

5.3. **Discussion.** Instructors account for roughly 60-65% of cross-sectional variation in political characteristics across courses, with the remainder attributable to environmental factors such as student demand, administrative preferences, and local governance. While the instructor component is the majority, instructors still respond to their environments, just like many economic actors.

What these "instructor effects" are is open for interpretation. Research-active faculty develop subject matter expertise that is portable when they move. Insofar as some topics are inherently political or carry a partisan valence, an instructor who continues teaching in their area of specialization will perpetuate that political character regardless of where they teach. While institutions exercise some discretion over which courses appear in the catalog and which instructors they hire, the results of this section suggest that the accumulated expertise of the faculty is a primary determinant of which courses are taught and how.

The finding that institutional context accounts for roughly 35% of cross-sectional variation is not trivial. Context can operate through several channels: promotion committees that shape faculty incentives, curriculum committees that determine which courses are eligible to be offered, and administrative or legislative mandates that restrict or expand the topics instructors may address. The magnitude of the context effect suggests such interventions may bite; however, our estimates also imply that the instructor component, which is less directly amenable to top-down policy, is the larger of the two. In the following section, we address a specific part of institutional context: student demand.

6. Student Demand

The movers design shows that instructors respond meaningfully to the environments in which they teach. This section studies one stakeholder within those environments directly: students. Student choices create a demand-side force on course offerings. Courses that better match student

[19]Because the heterogeneity analysis requires thinner slices of data, we pool all "pre" and "post" periods ±5 years around each move. We limit to on-diagonal entries of the transition matrix — e.g., estimated effects of moving from one high selectivity school to another high selectivity school.

preferences attract greater enrollment, creating a channel through which student preferences can influence the incentives facing instructors and departments. We recover these preferences from observed enrollment using a nested logit model of course choice. Our results show that demand for political content is positive and relatively stable, while preferences for liberal content grow during the 2010s, peak in 2020, and partially reverse thereafter. The results speak to the dynamic patterns in course supply documented in Section 4, showing alignment between trends in course offerings and student preferences for liberal courses for most of our sample period.

6.1. **Model.** We model course enrollment using a nested logit framework (Berry, 1994), building on a literature that models student course and major choices as human capital investments (e.g., Arcidiacono, 2004; Beffy et al., 2012; Wiswall and Zafar, 2015; Ahn et al., 2024). Each institution (i) by semester (t) is a market, with fields as nests. Students first choose a field and then choose a course section within the field; the outside option is choosing not to take a course (an option not in any nest). An idiosyncratic type-I extreme value taste shock $\varepsilon_{jt}(\sigma)$ generates the nested logit structure with nesting parameter σ. We compute each course section's choice probability p_{jt} as its share of total enrollment during a given semester, and within-nest choice probabilities $p_{j|f(j),i(j)t}$. The estimating equation is:

$$\begin{aligned} \ln(p_{jt}) &= \kappa_{i(j)t} + \kappa_{i(j)f(j)} + \kappa_{\text{level}(j)} + \kappa_{f(j)y(t)} + \sigma \ln p_{j|f(j),i(j)t} \\ &\quad + \sum_{s} \mathbb{1}\{s \in y(t)\} \times \left(\beta_{1s} \times \text{political content}_{jt} + \beta_{2s} \times \text{slant}_{jt}\right) + \xi_{jt} \end{aligned} \tag{6.1}$$

The coefficients of interest, β_{1s} and β_{2s}, are students' preferences for political content and slant. We allow these preferences to vary by (academic) year. We include fixed effects at the institution-by-semester ($\kappa_{i(j)t}$), field-by-school ($\kappa_{i(j)f(j)}$), course level $\kappa_{\text{level}(j)}$,[20] and field-by-year ($\kappa_{f(j)y(t)}$) level. The term $\sigma \ln(p_{j|f(j),i(j)t})$ captures the within-nest substitution pattern.[21] ξ_{jt} captures unobserved course characteristics that affect enrollment. We assume that within an institution-by-field-by-semester cell, ξ_{jt} is serially uncorrelated over time once the student body turns over, i.e. every four

[20]Determined based on each institution's course numbering protocol. For example, lower-level introductory courses are typically numbered in the 100-299 range; upper-level advanced undergraduate courses are typically numbered in the 300-499 range.

[21]As is standard, we instrument the inside-nest share with the log number of courses in the field during that semester, which isolates variation in the number of close substitutes across schools with more and fewer courses per field. Course offerings are highly persistent in the short run, limiting concerns that the total number of course sections offered are endogenous to the demand shocks for individual courses (Light, 2026).

years. We omit the outside option term $\ln(p_0)$ because the market fixed effect $\kappa_{i(j)t}$ absorbs the (market-constant) outside share. When estimating, we cluster standard errors at the institution level.

We make two key simplifying assumptions when modeling student demand. First, because we only observe aggregate market shares, we treat decisions at the seat level rather than at the individual level. Doing so abstracts from complementarities and conflicts in scheduling. It is convenient to think of these concerns as being captured in ε_{jt} — a student who would otherwise take course j but faces a conflict simply draws a low ε_{jt}. Second, we abstract away from course capacity constraints. Where caps bind, we observe a censored choice set and understate demand for the most sought-after courses and overstate demand for less sought-after courses. In practice, 76% of courses for which we observe caps have excess capacity, and this share is relatively stable throughout our sample.[22] Additionally, if students rationed out of high-demand courses substitute to courses with similar political characteristics, any bias should be minimal.

Conditional on the fixed effects, variation that allows us to estimate β_{1s} and β_{2s} comes from within field-by-year, within-market variation in enrollment across courses. The institution-by-semester fixed effects $\kappa_{i(j)t}$ absorb everything common to a market (e.g., a campus-wide enrollment shock). The institution-by-field fixed effect $\kappa_{i(j)f(j)}$ absorbs persistent cross-institution differences in field quality (e.g., a school's specialization in a field). The course level effect $\kappa_{\text{level}(j)}$ absorbs differences across introductory, advanced undergraduate, and graduate courses. Finally, the field-by-year effect $\kappa_{f(j)y(t)}$ absorbs any movement common to a field in a year, including secular shifts in field popularity (e.g., the rise of STEM) and, importantly, any aggregate changes in demand for political characteristics operating at the field level. Estimation of β_{1s} and β_{2s}, therefore, relies only on variation not absorbed by these fixed effects: whether the more political or slanted course in a given field, institution, and semester draws disproportionate enrollment.

6.2. **Identification.** The primary identification challenge is that course political characteristics and enrollment are jointly determined. Students sort toward courses that match their preferences, but instructors also adjust content in response to institutional contexts, which includes student demand

[22]Appendix D.3 summarizes diagnostic statistics on capacity-constrained courses. There is no clear relationship between a course's political content and the likelihood of enrolling to capacity; liberal slant is at most weakly correlated with the likelihood of enrolling to the cap. Enrolling to capacity is an extensive form of potential demand censoring; we are unable to say anything about whether the number of students rationed out of high-demand courses relates to course political characteristics.

(Section 5). Consequently, OLS recovers an equilibrium relationship between course content and enrollment, rather than student preferences.

This simultaneity induces endogeneity because instructors may alter political characteristics in response to unobserved demand shocks. For example, an instructor who expects unusually weak enrollment (low ξ_{jt}) may respond by adjusting political characteristics to make their course more attractive. Such behavior would create a correlation between ξ_{jt} and the characteristics students value, biasing the $\hat{\beta}$ estimates toward zero.[23] Furthermore, $\ln(p_{j|f(j),i(j)t})$ inherits these issues as well.

To break the simultaneity, we instrument for the political characteristics of an instructor's current courses using the average political characteristics of the same instructor's courses four years prior. The four-year lag ensures the student body in year t is largely disjoint from that in $t-4$, so cohort-specific demand shocks faced by the instructor in the earlier period will not carry forward to the current cohort. For multi-instructor courses, we take the average of the lags across instructors. Given that the number of section offerings is inelastic in the short run (Light, 2026), we use the log of the number of course offerings in a field ($\ln(n_{f(j)|i(j)t})$) as an instrument for $\ln(p_{j|f(j),i(j)t})$. The first stage is:

$$\begin{aligned}
\psi_{jt} = \alpha_0 &+ \kappa_{i(j)t} + \kappa_{i(j)f(j)} + \kappa_{\text{level}(j)} + \kappa_{f(j)y(t)} + \alpha_1 \ln(n_{f(j)|i(j)t}) \\
&+ \alpha_2 \frac{1}{|Instructor(j)|} \sum_{j' \in Inst(j)} \text{political content}_{j't-4} \\
&+ \alpha_3 \frac{1}{|Instructor(j)|} \sum_{j' \in Inst(j)} \text{slant}_{j't-4} \\
&+ \nu_{jt}
\end{aligned}$$

Substituting $\sum_s \mathbb{1}\{s \in y(t)\}\beta_{1s} \times \text{political content}_{jt}$ or $\sum_s \mathbb{1}\{s \in y(t)\}\beta_{2s} \times \text{slant}_{jt}$ for ψ_{jt} when instrumenting for political content and slant, respectively. Given the fixed effects and Frisch-Waugh-Lovell, the instrument isolates the instructor's persistent political deviation from their own field-year and institution-field. The intuition is that an instructor who taught noticeably to the left of their field four years ago is very likely teaching to the left of it today, but the specific cohort, salience,

[23]This parallels a standard product market dynamic: a producer of a low-quality product attempts to compensate by lowering the price to increase demand. In this case, ideological appeal served the role of the "price" aligning supply and demand.

and enrollment conditions that drove demand in their courses four years ago have turned over. Lagging therefore retains the portion of current content that is predictable from the instructor's stable political orientation while discarding the portion that co-moves with the current environment. The instrument is relevant because course offerings tend to be sticky: titles and learning objectives constrain revision, course revisions involve fixed instructor effort costs, and courses require approval from other university authorities, such that an instructor's political positioning in $t-4$ strongly predicts their positioning in t.

Exclusion requires that an instructor's lagged political positioning affects current enrollment only through its persistence into current course content. The four-year lag and fixed effects substantially reduce concern for three potential violations. First, the student body in t is largely disjoint from the student body in $t-4$, limiting persistence of cohort-specific demand shocks. Second, field-year fixed effects absorb demand shocks common to a field-year (e.g., a discipline-wide surge of interest in immigration or climate change). Third, institution-by-field fixed effects absorb stable within institution-by-field differences in course requirements.

The principal remaining threat to exclusion is that instructors' persistent political positioning may proxy for persistent instructor characteristics that independently affect enrollment. For example, instructors teaching more political courses may also be systematically more engaging lecturers, more lenient graders, or otherwise more desirable to students. Because these attributes are persistent, they appear both in the instructor's lagged courses and current courses.

We assess this concern using parallel course sections of the same course taught by different instructors. For example, when two instructors teach Economics 101 in the same institution and semester using the same catalog description, differences in enrollment across sections reveal differences in instructor desirability while holding the course's political characteristics fixed.[24] For each instructor, we use these enrollment differences to construct a measure of desirability and test whether it is correlated with the average political characteristics of the instructor's *other* courses.

[24] 56.0% of instructors teach at least one parallel section over our sample period. However, only 16.5% of courses offered include parallel sections.

Figure D.1 shows instructor desirability is uncorrelated with the political content or slant of their other courses.[25] This result provides evidence that the instrument is not proxying for a broader, content-independent dimension of instructor desirability correlated with political characteristics.

This exercise does not rule out every possible exclusion concern. Political content and slant may correlate with persistent course characteristics unrelated to instructor desirability, such as specialized human capital or labor market relevance. This will bite hardest in disciplines where political and skill content are bundled within a field — e.g., in Economics, if applied econometrics and public economics differ in both job-relevant skills and political characteristics. This concern is likely more important for political content than for slant. Conditional on political content, we argue slant carries little job-relevant skill content (a liberal and a conservative applied course are plausibly similar in human capital), so the orthogonality assumption is more defensible for β_{2s} than for β_{1s}.

The IV has the added benefit of partially addressing measurement error in political content and slant. The instrument is computed in units of political content and slant, though potentially from different courses. When an instructor's lagged courses are entirely disjoint from the current courses, our instrument removes attenuation as long as realizations of the measurement error are independent. In the case where an instructor teaches the same courses over an extended period of time, the same measurement error may affect both the realization of political characteristics of the regressor and the instrument. In the limit where instructors teach many distinct courses, the shared measurement error becomes small and attenuation falls to zero. Even when some identical course descriptions are shared across the four-year gap, this overlap is typically incomplete, meaning the IV estimate will still be attenuated toward zero by a lesser degree than OLS.

The IV requires we observe each instructor for four years, which excludes new hires. If schools respond to demand shifts partially by hiring new faculty with corresponding political profiles, the IV will underweight this margin and may attenuate the demand response. Applying the same sample

[25] Two validations of our instructor desirability measure ensure we are not regressing on noise. First, we perform a split-half reliability check: Figure D.2a shows that quality estimated using even years alone is strongly correlated with desirability estimated using only odd years, so the measure reflects a stable signal rather than sampling variation. Second, we perform an out-of-sample check using non-shared courses: Figure D.2b shows that instructors we estimate to be high quality based on their shared (multi-section) courses draw significantly higher enrollment in their single-instructor classes. The quality measure, therefore, captures a persistent and demand-relevant object not specific to politically slanted content.

restriction to our OLS estimation ensures quantities are comparable across specifications.[26] We estimate a strong first stage (joint $F = 251$; Table D.1).

We present our results as semi-elasticities of enrollment with respect to a one standard deviation change in political content or slant. Our nested logit implies that, given a coefficient $\hat{\beta}$ and nesting parameter $\hat{\sigma}$, the semi-elasticity for characteristic x with respect to enrollment is:

$$\frac{\partial \ln s_j}{\partial x} = \hat{\beta} \times \left(\frac{1 - p_{j|f(j),i(j)t}}{1 - \hat{\sigma}} + p_{j|f(j),i(j)t} - p_{jt} \right)$$

We average this quantity across all courses j within each year to compute the average annual semi-elasticity. We compute standard errors by clustering at the institution level and then applying the Delta Method.

6.3. **Results.** Figure 8 plots the average semi-elasticities based on OLS estimates of β_{1s} and β_{2s} for each year in our sample using the structure of the nested logit model. As such, the estimates show the expected percentage change in enrollment with respect to a one-standard-deviation change in political content or slant, respectively. Recall from Section 4 that, over the past 25 years, course offerings have moved only 0.15 SD more political, so this benchmark is much larger than the time trends we document in course offerings.

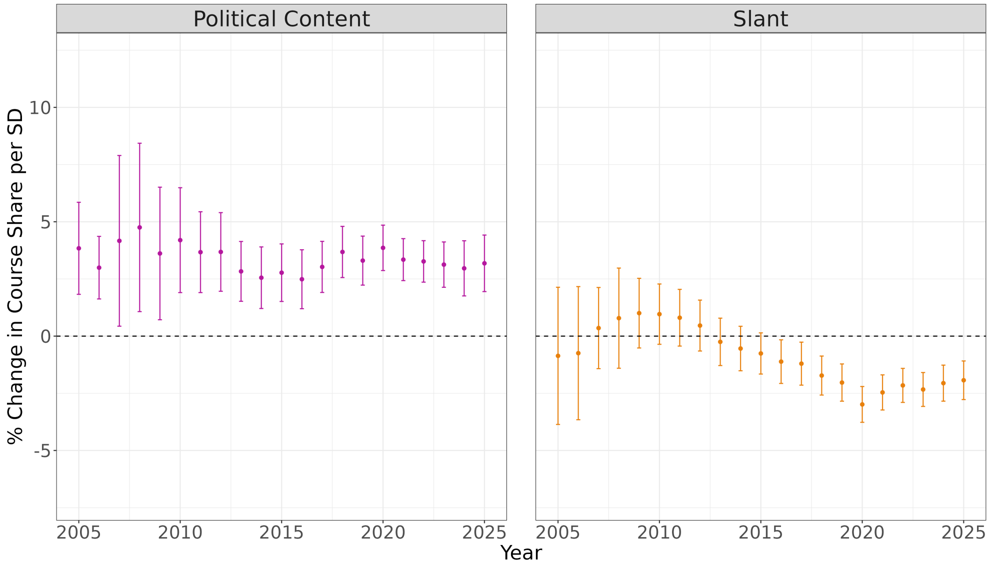
The enrollment response to political content is positive: a course one standard deviation more political would draw roughly 3.4 percent more enrollment relative to the mean course. The political content trend is relatively flat over time, with a relative peak in 2020 (albeit one we cannot statistically distinguish from the other coefficients).

The enrollment response to slant changes over time: from 2005 to the mid-2010s, enrollment did not respond to slant. Starting in the mid-2010s, however, students began to prefer liberal content with a distinctive peak in 2020. A course one standard deviation more liberal would have enrollment roughly 2.1 percent higher in 2020 than the same course in 2005. Nesting by field is empirically important for model fit; we estimate the nesting parameter $\sigma = 0.159$ (0.014), indicating there is more substitution on the margin of political content and slant *within* a field than *between* fields.

The semi-elasticity estimates based on the IV specification in Figure 9 recover larger magnitudes and trends than those based on OLS. A course one standard deviation more political would draw roughly 5 percent more enrollment (in an average year) relative to the mean course. While there

[26]Note this marginally changes the estimand — rather than capturing demand for political content generally, we capture demand for political content among established instructors.

FIGURE 8. OLS semi-elasticities of student demand for political content and slant

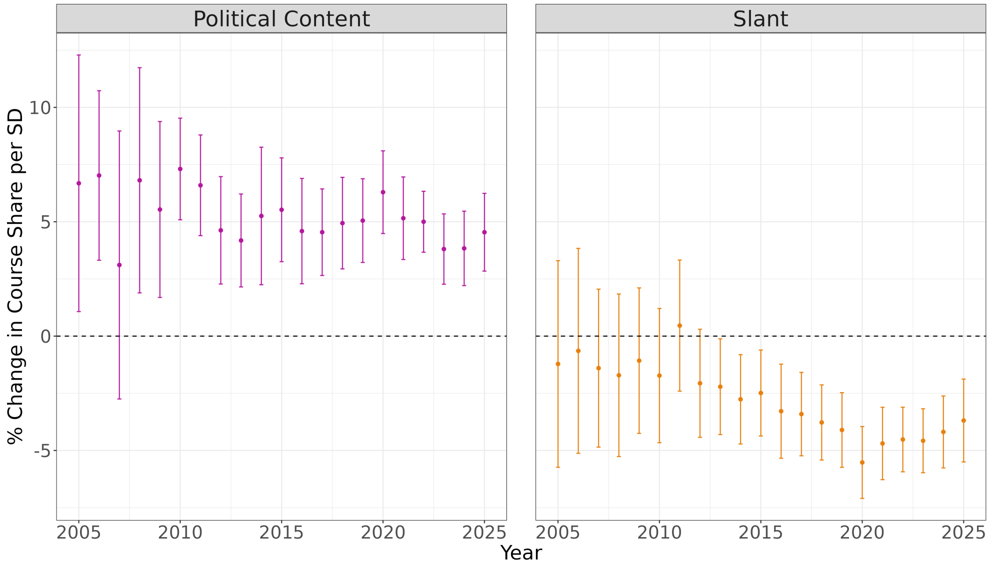


Notes: Points are nested-logit semi-elasticities based on OLS coefficients from a regression of political content and slant by year from the nested logit demand system in Equation 6.1. Values are the semi-elasticity of course enrollment with respect to a one-standard-deviation increase; negative slant coefficients indicate a preference for more liberal content. Standard errors are computed via the Delta Method from OLS standard errors clustered at the institution level.

is still no preference for liberal content at the start of our period, by 2020 a course one standard deviation more liberal would see 5.5 percent more demand — almost doubling relative to the baseline OLS estimate of 3 percent. Relative to the surrounding years, 2020 sees peak demand for both political content and liberal slant. Unlike the smooth descriptive trends, both coefficients exhibit a spike during the 2020-2021 school year. Since then, preferences have started to reverse toward less political, less liberal courses. The nesting parameter estimate is essentially unchanged, $\sigma = 0.158$ (0.014).

Contrasting the estimates suggests that the OLS estimates are attenuated. This may be because instructors endogenously adjust content toward student preferences when they have a particularly low ξ_{jt} draw; for example, an instructor teaching a relatively narrow or specialized course may emphasize more broadly appealing political topics in an effort to attract enrollment. Our instrument may also correct for measurement error, which would also attenuate the estimates based on OLS coefficients.

FIGURE 9. IV semi-elasticities of student demand for political content and slant



Notes: Points are nested-logit semi-elasticities based on IV coefficients from a regression of political content and slant by year from the nested logit demand system in Equation 6.1, instrumenting each course's current characteristics with the average characteristics of courses taught by the same instructor four years earlier. Values are the semi-elasticity of course enrollment with respect to a one-standard-deviation increase; negative slant coefficients indicate a preference for more liberal content. Standard errors are computed via the Delta Method from IV standard errors clustered at the institution level.

6.4. **Discussion.** Our estimates suggest that student demand for political courses has been positive since 2005; however, growing student preference for liberal courses is a more recent phenomenon. These dynamics complicate any narrative in which universities unilaterally shape student political exposure: the trend toward more liberal course offerings aligns with contemporaneous increases in demand for liberal content. The substantial movement in slant semi-elasticities across our sample period indicates that student demand is not static, and changes in offerings occur in a broader context of market preferences.

The relationship between course offerings and enrollment over our sample period admits multiple interpretations. In the early part of our sample, courses were becoming more liberal on average, but students were not particularly responsive in terms of choosing which courses to enroll in. Later in the sample, as student preferences for liberal content increased, universities continued to offer more liberal content. This pattern is consistent with universities responding to changing student preferences over the second half of our sample. We caution against interpreting this trend as universities influencing student preferences — we have no evidence for or against the sources of student preference changes.

7. Conclusion and Future Work

This paper provides systematic evidence on the political characteristics of U.S. college instruction and the forces that shape it. We develop a text-based method that separately measures the amount of political content in short texts and the partisan slant of this content. Applying this method across nearly 1,000 colleges and universities over 25 years, we show that courses have become modestly more political and more liberal. These changes are precisely estimated, but modest — political content changes are equivalent to roughly one-quarter of the difference between the average Psychology and History course, and the shift in slant is equivalent to roughly one-fifth of the difference between the slant of the average Economics and Sociology course — suggesting that differences in students' political exposure across fields remain substantially larger than recent changes over time.

We also provide evidence on how instructors shape these course characteristics. Instructors who move across institutions adjust their courses toward their destination, closing 35-40% of the political content and slant gaps between their origin and destination programs, implying that instructors contribute 60-65% of variation in political characteristics of courses. The substantial persistence of instructors' course content indicates that instructor-specific factors travel across institutional settings, while leaving some scope for the importance of local contexts.

On the demand side, students exhibit positive demand for political content throughout our sample, while preferences for liberal content emerge during the 2010s, peak in 2020, and partially reverse thereafter. These results suggest that the political characteristics of college courses reflect the interaction of instructors, students, and institutions, rather than the preferences of any single group.

Our analysis is deliberately positive, rather than normative. Slant measures whether course content more closely resembles one partisan pole or the other; it does not measure truth or ideological *bias*. Liberal slant could reflect ideological advocacy, differences in topics emphasized by the parties, or evidence that aligns more closely with one party's position. For example, the scientific consensus on anthropogenic climate change (Oreskes, 2004) is reflected in new courses that, relative to the party platforms, code more liberal; analogous alignments in other directions are possible on other topics. Similarly, student preferences do not provide a welfare criterion: universities may value exposure of students to perspectives or knowledge they would not independently choose. Therefore, we interpret our estimates as describing *how* political content is produced and selected, not whether the resulting equilibrium is *desirable*.

Several questions remain unanswered. Course descriptions represent only one channel through which students experience politics in college; measuring political content in course instruction and on campus more broadly would complement our analysis. Further, linking exposure to political content to downstream outcomes, such as earnings or voting behavior, could speak to more welfare-relevant questions (e.g., Goldstein and Kolerman, 2025; Baum et al., 2026). We do not model substitution between universities or inter-university competition, though politics seems to influence student choices on these margins as well (Acton et al., 2024). Finally, we say little about university governance — the political economy of academic bureaucracy appears to be a fruitful avenue for future research.

References

Acton, R., E. Cook, and P. Ugalde Araya (2024): "Political Views & College Choices in a Polarized America," Tech. rep., Working Paper.

Ahn, T., P. Arcidiacono, A. Hopson, and J. Thomas (2024): "Equilibrium grading policies with implications for female interest in STEM courses," *Econometrica*, 92, 849–880.

Andrews, M. J., W. Marble, and L. Russell (2025): "The Long-Run Effects of Colleges on Civic and Political Life," Tech. rep., Center for Open Science.

Arcidiacono, P. (2004): "Ability sorting and the returns to college major," *Journal of econometrics*, 121, 343–375.

Baum, M., J. Endara, and A. Paulson (2026): "Do Faculty Affect Student Partisanship?" Tech. rep., Working Paper.

Beffy, M., D. Fougere, and A. Maurel (2012): "Choosing the field of study in postsecondary education: Do expected earnings matter?" *Review of Economics and Statistics*, 94, 334–347.

Bell, D., J. B. Holbein, S. J. Imlay, and J. Smith (2024): "Which Colleges Increase Voting Rates?" Tech. rep., IZA Discussion Papers.

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 242–262.

Bloom, A. (1987): "The Closing of the American Mind: How Higher Education Has Failed Democracy and Impoverished the Souls of Today's Students," New York: Simon & Schuster.

Bolukbasi, T., K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai (2016): "Man is to computer programmer as woman is to homemaker? debiasing word embeddings," *Advances in neural information processing systems*, 29.

Boxell, L. and J. Conway (2022): "Journalist ideology and the production of news: Evidence from movers," *Available at SSRN 4295587*.

Braghieri, L., S. Eichmeyer, R. Levy, M. Mobius, J. Steinhardt, and R. Zhong (2024): "Article level slant and polarization of news consumption on social media," *Available at SSRN*, 4932600.

Buckley, Jr., W. F. (1951): *God and Man at Yale: The Superstitions of "Academic Freedom"*, Chicago: Henry Regnery Company.

Cagé, J., M. Hengel, N. Hervé, and C. Urvoy (2025): "Political Bias in the Media–Evidence from the Universe of French Broadcasts, 2002-2020," Tech. rep., CESifo Working Paper.

Cantoni, D., Y. Chen, D. Y. Yang, N. Yuchtman, and Y. J. Zhang (2017): "Curriculum and ideology," *Journal of political economy*, 125, 338–392.

Chen, W.-L., M.-J. Lin, and T.-T. Yang (2023): "Curriculum and national identity: Evidence from the 1997 curriculum reform in Taiwan," *Journal of Development Economics*, 163, 103078.

Chen, Y., Y. Chi, J. Fan, and C. Ma (2021): "Spectral methods for data science: A statistical perspective," *Foundations and Trends in Machine Learning*, 14, 566–806.

Compiani, G., I. Morozov, and S. Seiler (2026): "Demand estimation with text and image data," *The RAND Journal of Economics*.

Dee, T. S. (2004): "Are there civic returns to education?" *Journal of public economics*, 88, 1697–1720.

Dell, M. (2025): "Deep learning for economists," *Journal of Economic Literature*, 63, 5–58.

Ethayarajh, K., D. Duvenaud, and G. Hirst (2019): "Understanding undesirable word embedding associations," *arXiv preprint arXiv:1908.06361*.

Finkelstein, A., M. Gentzkow, and H. Williams (2016): "Sources of Geographic Variation in Health Care: Evidence From Patient Migration*," *The Quarterly Journal of Economics*, 131, 1681–1726.

Firoozi, D. (2025): "Education and Partisanship," Tech. rep., Working Paper.

Garg, N., L. Schiebinger, D. Jurafsky, and J. Zou (2018): "Word embeddings quantify 100 years of gender and ethnic stereotypes," *Proceedings of the National Academy of Sciences*, 115, E3635–E3644.

Gentzkow, M., B. Kelly, and M. Taddy (2019a): "Text as data," *Journal of Economic Literature*, 57, 535–574.

Gentzkow, M. and J. M. Shapiro (2010): "What drives media slant? Evidence from US daily newspapers," *Econometrica*, 78, 35–71.

Gentzkow, M., J. M. Shapiro, and M. Taddy (2019b): "Measuring group differences in high-dimensional choices: method and application to congressional speech," *Econometrica*, 87, 1307–1340.

Goldstein, Y. and M. Kolerman (2025): "Changing Minds: How Academic Fields Shape Political Attitudes," SSRN Scholarly Paper 5196889, Working Paper.

Groseclose, T. and J. Milyo (2005): "A measure of media bias," *The quarterly journal of economics*, 120, 1191–1237.

Hitchcock, J. (1971): "Colleges: No One Has a Good Word For Them," *The New York Times*.

Huh, M., B. Cheung, T. Wang, and P. Isola (2024): "The platonic representation hypothesis," *arXiv preprint arXiv:2405.07987*.

Jelveh, Z., B. Kogut, and S. Naidu (2024): "Political language in economics," *The Economic Journal*.

Kozlowski, A. C., M. Taddy, and J. A. Evans (2019): "The geometry of culture: Analyzing the meanings of class through word embeddings," *American Sociological Review*, 84, 905–949.

Light, J. (2026): "Student demand and the supply of college courses," *Available at SSRN 4856488*.

Longuet-Marx, N. (2024): "Party Lines or Voter Preferences? Explaining Political Realignment," Tech. rep., Working Paper.

Magnolfi, L., J. McClure, and A. Sorensen (2025): "Triplet embeddings for demand estimation," *American Economic Journal: Microeconomics*, 17, 282–307.

Martin, G. J. and A. Yurukoglu (2017): "Bias in cable news: Persuasion and polarization," *American Economic Review*, 107, 2565–2599.

Neelakantan, A., T. Xu, R. Puri, A. Radford, J. M. Han, J. Tworek, Q. Yuan, N. Tezak, J. W. Kim, C. Hallacy, et al. (2022): "Text and code embeddings by contrastive pre-training," *arXiv preprint arXiv:2201.10005*.

Oreskes, N. (2004): "The scientific consensus on climate change," *Science*, 306, 1686–1686.

Park, K., Y. J. Choe, and V. Veitch (2023): "The linear representation hypothesis and the geometry of large language models," *arXiv preprint arXiv:2311.03658*.

Shields, J. A., Y. Avnur, and S. Muravchik (2025): "Closed classrooms? An analysis of college syllabi on contentious issues," *Unpublished paper, Claremont-McKenna College, Department of Government*.

Vershynin, R. (2018): *High-dimensional probability: An introduction with applications in data science*, vol. 47, Cambridge university press.

Widmer, P., S. Galletta, and E. Ash (2022): "Media slant is contagious," *arXiv preprint arXiv:2202.07269*.

Wiswall, M. and B. Zafar (2015): "Determinants of college major choice: Identification using an information experiment," *The Review of Economic Studies*, 82, 791–824.

Yu, Y., T. Wang, and R. J. Samworth (2015): "A useful variant of the Davis–Kahan theorem for statisticians," *Biometrika*, 102, 315–323.

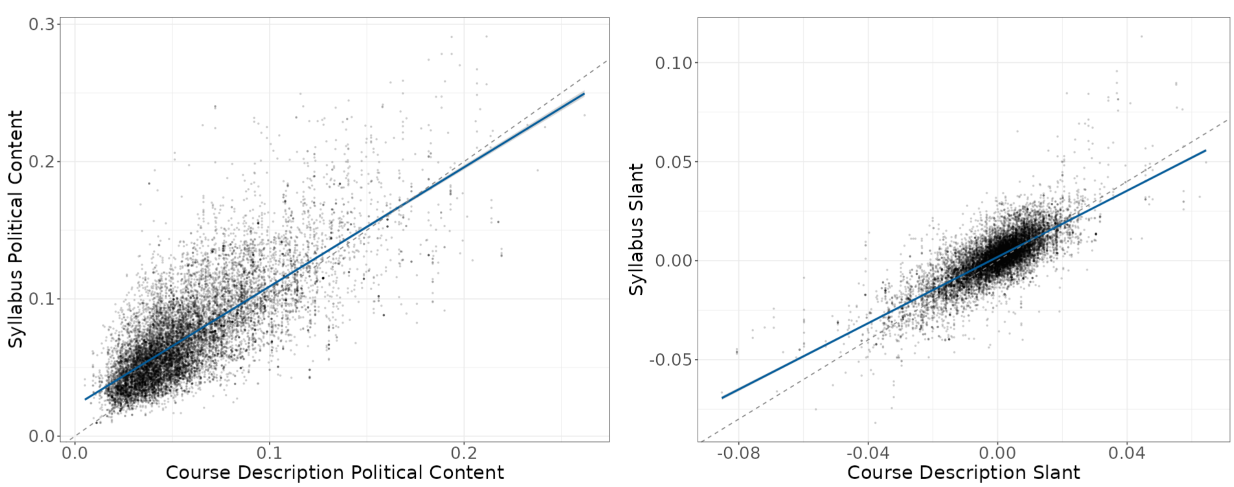
Zingher, J. N. (2022): "TRENDS: Diploma divide: Educational attainment and the realignment of the American electorate," *Political Research Quarterly*, 75, 263–277.

Appendix A. Measurement Supplement

This appendix contains supplementary material for Section 3, focusing on properties and validation of the projection-based slant measurement.

A.1. **Supplementary Exhibits.**

Figure A.1. Description-based versus syllabus-based measures

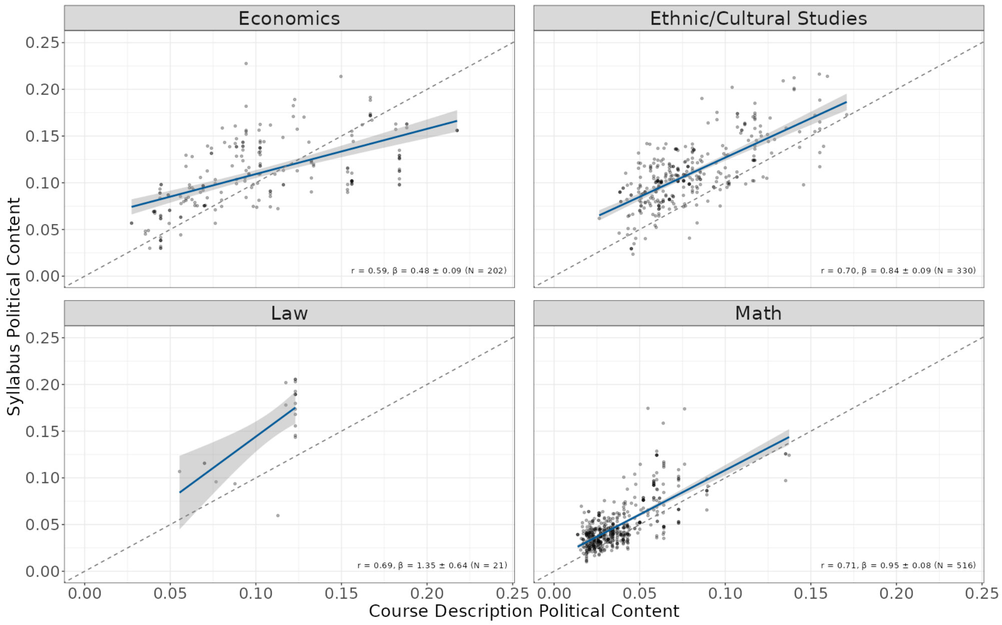


(a) Political content

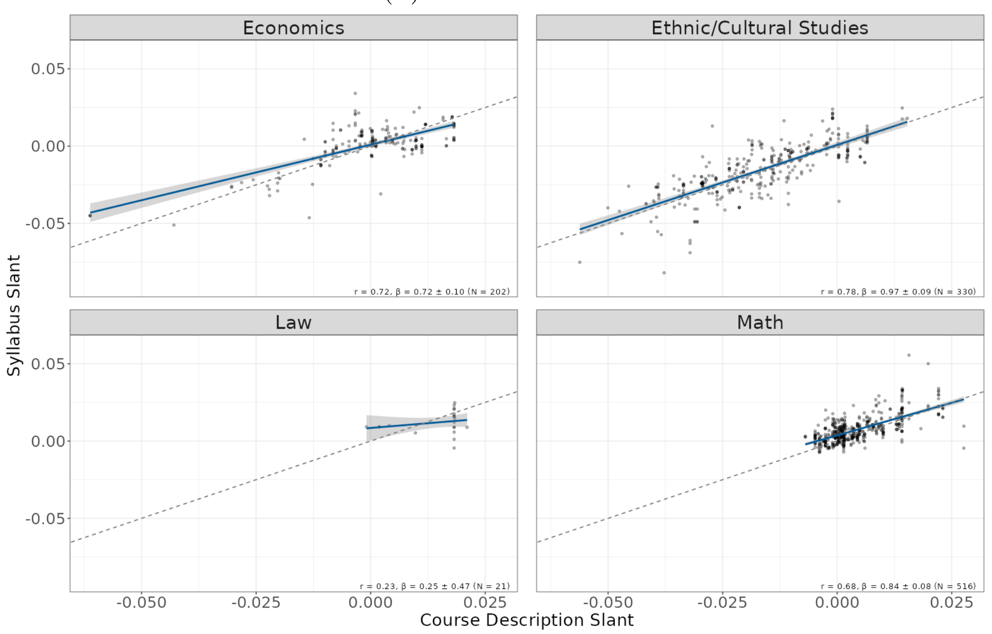
(b) Slant

Notes: Each panel compares the measure computed from a catalog course description (horizontal axis) with the same measure computed from a matched syllabus for courses at a large public university. Panel (a) reports political content and Panel (b) slant. Both description- and syllabus-based measures are constructed using the same 2016 party platform benchmarks. The correlations between description- and syllabus-based measures are 0.77 for political content and 0.78 for slant. The 45-degree line denotes equality between the two measures, so proximity to the line indicates agreement in cardinal values rather than only rank.

FIGURE A.2. Description-based versus syllabus-based measures, split by field



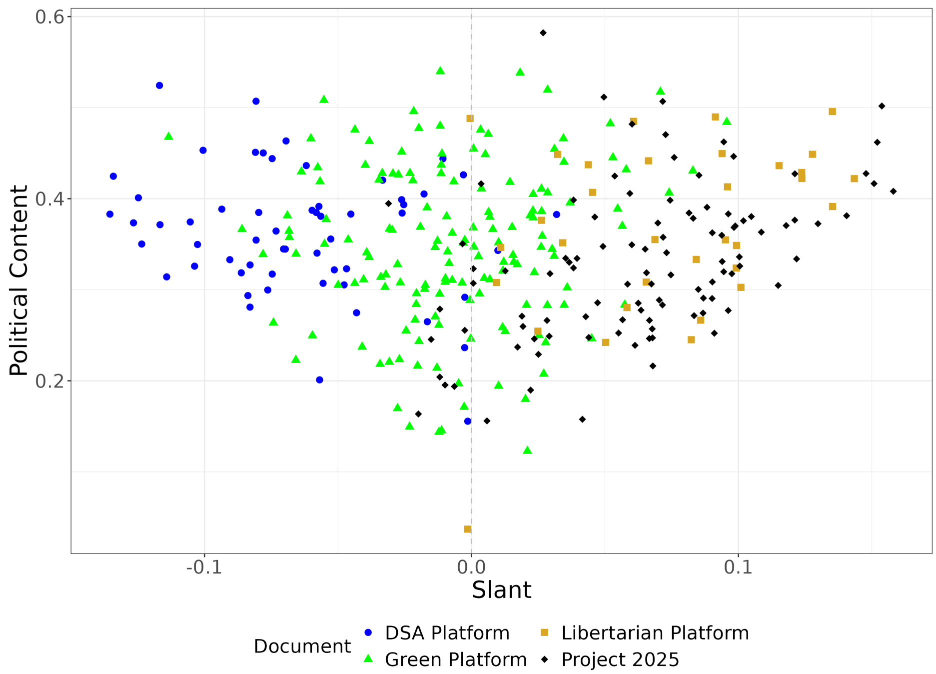
(A) Political content



(B) Slant

Notes: Each panel compares measures constructed from catalog course descriptions (horizontal axis) with measures constructed from matched syllabi (vertical axis), separately for the fields shown in the figure at a large public university. Panel (a) reports political content and Panel (b) slant. Both measures use the same 2016 party-platform benchmarks. The 45-degree line denotes equality between the two measures, so proximity to the line indicates agreement in cardinal values rather than only rank.

FIGURE A.3. Out-of-sample political content and slant of known political texts



Notes: The figure applies the political content and slant measures to sections of political texts that are not used to construct the baseline 2016 Democratic and Republican platform benchmarks. Each point represents one paragraph or section of the indicated document. Scales are comparable to those in the course data.

A.2. **Econometric Details.** Let c denote a generic course representation vector. For a matrix X, let P_X denote the matrix that projects onto the span of the columns of X. We can begin by re-writing the definitions of political content and slant in terms of projection matrices. For the sake of concreteness, we set R to be the representation vectors of the conservative platform (instead of the generic C used in the main text). The matrix D collects the same for the Democratic platform. Then, define $X := [R|D]$ where $|$ denotes horizontal concatenation. As a useful property of the inner product, we can redefine political content as a projection:

$$p(c|R, D) = \cos^2(\theta_{cX}) = \cos^2(\theta_{cX})||c||^2 = \|P_X c\|^2 = \langle P_X c, P_X c\rangle = \langle P_X^T P_X c, c\rangle = \langle P_X c, c\rangle$$

Using the fact c is a unit vector, standard geometric properties of projection, along with the idempotency and symmetry of the projection matrix, we can prove the following numerical equivalence:

Proposition 1. *Assume $p(c|R, D) > 0$. Then the following are numerically equivalent:*

$$s(c|R, D) = \cos^2(\theta_{cR}) - \cos^2(\theta_{cD})$$

$$= p(c|R, D) \cdot v(c|R, D)$$

Where $p(c|R, D)$ is the political content of c and $v(c|R, D)$ is the valence of c.

Proof. We can re-write slant:

$$\begin{aligned}
s(c|R, D) &= \cos^2(\theta_{cR}) - \cos^2(\theta_{cD}) \\
&= \|P_R c\|^2 - \|P_D c\|^2 \\
&= \|P_R P_X c\|^2 - \|P_D P_X c\|^2 \\
&= \|P_R c^\perp\|^2 - \|P_D c^\perp\|^2 \\
&= \|c^\perp\|^2 \left(\frac{\|P_R c^\perp\|^2}{\|c^\perp\|^2} - \frac{\|P_D c^\perp\|^2}{\|c^\perp\|^2} \right) \\
&= \cos^2(\theta_{cX}) \left[\cos^2(\theta_{c^\perp R}) - \cos^2(\theta_{c^\perp D}) \right] \\
&= p(c|R, D) \cdot v(c|R, D)
\end{aligned}$$

Where the first line follows from the same step as when manipulating the definition of political content and the definition of the inner product. Then, the second step (inserting P_X) follows from noting the span of R and span of D are both contained in the span of X. The third step follows from $P_X c = c^\perp$ by definition, as $c^\perp$ is the orthogonal projection onto the span of X. The fourth step follows from multiplying and dividing by $\|c^\perp\|^2$. The fifth step follows from noting that $\cos^2(\theta_{vH})$ for vector v and subspace H is equal to $\|P_H v\|^2/\|v\|^2$, using the definitions of orthogonal projection and cosines. We apply this identity to each portion of the expression to get the sixth line. The last line follows from definitions. □

Valence can be computed empirically by dividing slant by political content; given that the measurement involves a quotient, it is correspondingly less well-behaved econometrically. We therefore focus on the properties of political content and slant.

This measure has several appealing properties. Geometrically, $\cos^2(\theta_{zX})$ is the squared magnitude of the projection of z onto $\mathcal{X}$: it equals 1 if the course lies entirely within the political subspace and 0 if the course is orthogonal to it. Statistically, it is equivalent to the non-centered R^2 from regressing the elements of z on the basis vectors spanning $\mathcal{X}$ (without an intercept). Intuitively, a course whose content is better "explained" by political language scores as more political. Finally, the measure is robust to representation noise. We give conditions under which rotational error in the representation vectors yields classical measurement errors in political content and slant.

A.2.1. *Measurement Error in the Course Vectors.* Our construction of the political content and slant measures ensures that measurement error in the course embeddings remains tractable in our main measures. We consider the case of rotational measurement error in the representations: given that all vectors we consider are unit length, we focus on errors in the direction, rather than magnitude, of the representations. The following section gives conditions under which the resulting measurement error for political content and slant will be classical.

Assume that we observe unit vector $\tilde{c}$, while the true latent representation is unit vector c. Define $a = c'\tilde{c}$, a scalar value that measures how close the observed representation $\tilde{c}$ is to the true representation c. Then, define $\eta = (1 - cc')\tilde{c}$, which is the portion of $\tilde{c}$ that is orthogonal to c. Then, we have that:

$$\tilde{c} = ac + \eta$$

We then have two useful identities: first, $c'\eta = 0$ by the orthogonality between c and η. Second, $a^2 + ||\eta||^2 = 1$ by the fact $\tilde{c}$ is a unit vector.

Assumption 1. *Assume that we observe $\tilde{c}_i = a_i c_i + \eta_i$, where c_i is the true course representation vector, while a_i and η_i are as defined above. Let Ω_i include a set of observables that we condition on: including field, institution, and time fixed effects, along with information on description length. We make the following two assumptions on η_i:*

(1) *Representation error is centered:* $\mathbb{E}[a_i \eta_i \mid c_i, \Omega_i] = 0$

(2) *The quadratic forms of the orthogonal error component satisfy:*

$$\mathbb{E}[\eta_i'(P_X - p(c_i)I)\eta_i \mid c_i, \Omega_i] = \kappa_p$$

$$\mathbb{E}[\eta_i'(P_R - P_D - s(c_i)I)\eta_i \mid c_i, \Omega_i] = \kappa_s$$

For constants κ_p, κ_s.

The first condition is a mean-zero condition on the (signed) direction of the representation error, the analogous condition of mean zero error in an additive measurement error framework. This condition requires that conditional on the true representation, the orthogonal component of the error does not have a systematic signed direction. This condition does not impose that all error directions must be equally likely. A sufficient, although stronger than necessary, condition is conditional antipodal symmetry: (a_i, η_i) and $(a_i, -\eta_i)$ have the same distribution conditional on c_i and Ω_i. Therefore, this condition can be read as imposing a form of symmetry on the measurement error.

The second condition is a quadratic bias condition of the error relevant for political content and slant. To further interpret the conditions, we can re-write the condition for political content as:

$$\begin{aligned}\kappa_p &= \mathbb{E}[\eta_i'(P_X - p(c_i)I)\eta_i \mid c_i, \Omega_i] \\ &= \mathbb{E}\big[||P_X\eta_i||^2 \mid c_i, \Omega_i\big] - p(c_i)\mathbb{E}\big[||\eta_i||^2 \mid c_i, \Omega_i\big]\end{aligned}$$

The first term is the expected squared magnitude of the error that falls in the political subspace, i.e. the expected political content of the error. The second term accounts for the amount of political content from the true representation as the vector with error is normalized to unit length. Rearranging terms, consider $\mathbb{E}\big[||P_X\eta_i||^2 \mid c_i, \Omega_i\big] = p(c_i)\mathbb{E}\big[||\eta_i||^2 \mid c_i, \Omega_i\big] + \kappa_p$. In the case where $\kappa_p = 0$, the fraction of error in the political subspace relative to the total covariance in the error must be in proportion to the amount of political content. In the case where $\kappa_p \neq 0$, the assumption allows for a systematic, constant shift of error into (or out of) the political subspace from the error, compared to the true political content implied by the representation.

Under these assumptions, we can show that resulting measurement error in political content and slant takes the form of a constant plus conditionally mean-zero noise. In the case where political content or slant is included as an outcome variable, Proposition 2 implies that measurement error will be classical. In the case where political content or slant is included as a regressor, the standard classical errors-in-variables results additionally require that the measurement error is uncorrelated with the error term in the outcome equation.

Proposition 2. *Under Assumption 1, measurement error from using political content or slant in a regression is a constant plus conditional mean-zero noise.*

Proof. Consider using $\tilde{c}_i$ to compute political content and expand the definition:

$$\begin{aligned}p(\tilde{c}_i) &= \langle P_X\tilde{c}_i, \tilde{c}_i\rangle \\ &= \langle P_X(a_ic_i + \eta_i), (a_ic_i + \eta_i)\rangle \\ &= a_i^2\langle P_Xc_i, c_i\rangle + 2a_ic_i'P_X\eta_i + \eta_i'P_X\eta_i\end{aligned}$$

Note that $a_i^2 = 1 - ||\eta_i||^2$, so we have:

$$p(\tilde{c}_i) = p(c_i) + 2a_i c_i' P_X \eta_i + \eta_i'(P_X - p(c_i)I)\eta_i$$

By part one of Assumption 1, the second term will be mean zero. By part two of Assumption 1, the last term will have constant expectation κ_p. Define $u_i^p = p(\tilde{c}_i) - p(c_i) - \kappa_p$, noting that $\mathbb{E}[u_i^p \mid c_i, \Omega_i] = 0$ by re-arrangement of the prior expression. Therefore, we can write:

$$p(\tilde{c}_i) = p(c_i) + \kappa_p + u_i^p$$

Similar analysis can be done for slant. □

It should be noted that the quadratic bias condition can be relaxed. First, consider the case where the quadratic form can vary with field, i.e. the expectation of the political content quadratic form is equal to $\kappa_{p,f(i)}$ for political content and the slant quadratic form is $\kappa_{s,f(i)}$ in expectation. The constants by field are left unrestricted – as long as field fixed effects are included in the regression, these constants are absorbed.

A.2.2. *Measurement Error in the Party Platforms.* Analysis of measurement error in the party platforms is substantially more involved. Recall the goal: to recover the space spanned by the representations of the political platforms from both parties, so that we can project onto it. Projection onto the columns of a matrix X is not generally linear in the columns of X. Here, we consider an asymptotic framework and utilize SVD, under which the observed platform embeddings converge to the subspace spanned by their latent representations. For each n, we consider n latent platform vectors of fixed dimension d observed with random error in the representation. As n increases, the array contains additional benchmark vectors; substantively, this corresponds to the thought experiment of having access to additional sections of the platforms written by the parties. Our goal is to show convergence of the projection onto the observed representations to the true, latent representations, which then implies consistency of political content and slant. In all these results, we take the rank of the matrix as given and known – we leave analysis of the case where the rank is estimated to future work. When we state norms on matrices, unless otherwise specified, we mean the operator norm.

We first define some notation. We work with a true matrix M_n, built out of the horizontally concatenated party platform representation vectors, with representative element m_j. Each column of the $d \times n$ matrix is unit normalized. All quantities indexed by j, a specific column, may also depend on n but we suppress this dependence for the sake of simpler notation. We assume that for all sufficiently large n, M_n has rank $r < d$ with $\mathcal{M}$ as the latent r-dimensional column space of M_n we would like to project onto. Denote the corresponding projection matrix onto $\mathcal{M}$ as P_M, and the projection matrix onto the orthogonal complement of $\mathcal{M}$ as $P_M^{\perp} = I_d - P_M$. We take the sequence of M_n as a deterministic sequence with a column space that stabilizes for sufficiently large n.

Similar to the analysis of measurement error in the course representation vectors, we consider rotational measurement errors for the platform representation vectors as well. We observe $\tilde{M}_n$, a matrix horizontally concatenating the observed, unit normalized representation vectors with representative element $\tilde{m}_j$. We define similar notation as in the course representation case: let $a_j := m_j'\tilde{m}_j$ and $\eta_j := (I_d - m_j m_j')\tilde{m}_j$. Then, by orthogonal decomposition, we have that:

$$\tilde{m}_j = a_j m_j + \eta_j$$

By construction, we have that $m_j'\eta_j = 0$ (as η_j is the portion of $\tilde{m}_j$ orthogonal to m_j) and that $a_j^2 + ||\eta_j||^2 = 1$.

Our empirical procedure applies a singular value decomposition directly on $\tilde{M}_n$. However, theoretically, we analyze a transformation of the matrix; our first lemma establishes exact equivalence of studying the left singular vectors of the platform matrix to the matrix we analyze here.

Lemma 3. *Let A be a $d \times n$ matrix. Let $r \leq d$. Let $u_1, ..., u_r$ be the left singular vectors of A corresponding to its r largest singular values. Then $u_1, ..., u_r$ are the left eigenvectors of $\frac{1}{n}AA'$ corresponding to the r largest eigenvalues. Therefore, the span of any r leading left singular vectors of A is equivalent to span of some r leading eigenvectors of $\frac{1}{n}AA'$, and vice versa.*

Proof. Let $A = U\Sigma V'$ be a singular value decomposition of A. Then, we have that $\frac{1}{n}AA' = \frac{1}{n}U\Sigma V'V\Sigma'U' = U\frac{\Sigma\Sigma'}{n}U'$ by the orthogonality of V. The matrix $\Sigma\Sigma'$ is diagonal, with diagonal entries equal to the squared singular values of A, so this is an eigendecomposition of $\frac{1}{n}AA'$. The columns of U are its eigenvectors and the corresponding eigenvalues are $\sigma_k(A)^2/n$. Because x^2/n is strictly increasing on the non-negative singular values, the ordering of the singular values of A is the ordering of the eigenvalues of $\frac{1}{n}AA'$, and the r largest singular values correspond exactly to the r largest eigenvalues. $\square$

Lemma 3 does not require A to have rank r, and we apply it to $\tilde{M}_n$, which will generically have full rank d. In the special case $A = M_n$, which has rank r, the r leading singular values are the only non-zero ones. Because the column space of a matrix is spanned by the left singular vectors corresponding to its non-zero singular values, the lemma implies that the leading r-dimensional eigenspace of $\frac{1}{n}M_nM_n'$ is exactly $\mathcal{M}$. The remainder of this section demonstrates that the leading r-dimensional eigenspace of $\hat{S}_n$, and therefore the leading r-dimensional left singular subspace of $\tilde{M}_n$, is close to $\mathcal{M}$.

We therefore analyze the following transformation of the observed representation matrix:

$$\hat{S}_n := \frac{1}{n}\tilde{M}_n\tilde{M}_n' = \frac{1}{n}\sum_{j=1}^{n}\tilde{m}_j\tilde{m}_j'$$

Then, $\hat{P}_M$ denotes the projection onto the leading r eigenvectors of $\hat{S}_n$. By Lemma 3, this projection matrix will be exactly the projection onto the leading r left singular vectors of $\tilde{M}_n$, which is our overall goal. We refer to $\hat{S}_n$ as the empirical scatter matrix, or the scatter matrix, as the values are a slightly scattered version of the true representation vectors.

Demonstrating convergence of projection onto $\tilde{M}_n$ to projection onto the true, latent representation requires three intermediate steps. We outline the strategy informally here, and state technical conditions precisely as they become relevant. Throughout, we analyze $\hat{S}_n$ instead of $\tilde{M}_n$ as it will be easier to work with, applying Lemma 3. We define two auxiliary objects: first, define $S_n^* = \mathbb{E}[\hat{S}_n \mid M_n]$, the mean of the transformation of the observed representation matrix. This expected representation may slightly rotate the latent space relative to the true value $\mathcal{M}$. Second, we define S_n^0, a version of S_n^* with the blocks that map vectors from $\mathcal{M}$ to $\mathcal{M}^\perp$ removed in order to characterize this rotation (we give a formal definition before Lemma 5). The major step of the proof will be to make each term in the following expression small:

$$\hat{S}_n - S_n^0 = \underbrace{\hat{S}_n - S_n^*}_{\text{Random fluctuation}} + \underbrace{S_n^* - S_n^0}_{\text{Deterministic Rotation}}$$

The first term is controlled by standard matrix concentration results. We control the second term by an assumption on the geometry of the error. This result demonstrates that $\hat{S}_n$ is close to S_n^0. Then, under an additional condition that the platform has sufficiently strong signal, we show that the leading eigenspace of S_n^0 is equal to $\mathcal{M}$. With these results, we show that projection onto the eigenspace of $\hat{S}_n$ converges to projecting onto the true latent sub-space $\mathcal{M}$. These results allow us to analyze political content and slant by applying the projection results to show consistency.

We impose the following structure on the measurement error in the observed platform vectors.

Assumption 2. *For each n:*

(1) *For every j, $\mathbb{E}[a_j \eta_j \mid M_n] = 0$.*
(2) *Conditional on M_n, the observed representations $\tilde{m}_1, ..., \tilde{m}_n$ are mutually independent.*
(3) *Define $\Gamma_j := \mathbb{E}[\eta_j \eta_j' \mid M_n]$, the conditional second moment matrix of the representation error. Let $\|.\|$ denote the operator norm. Then:*

$$\left\| P_M^{\perp} \left(\frac{1}{n} \sum_{j=1}^{n} \Gamma_j \right) P_M \right\| = o(1)$$

Define $\bar{\Gamma}_n = \frac{1}{n} \sum_{j=1}^{n} \Gamma_j$ for the sake of compact notation.

The first part of Assumption 2 is a centering condition, identical to our assumption in the case of error in the course representations (Assumption 1). Conditional on the true representations, the error must not have a systematic signed direction. The second part of Assumption 2 imposes independence in the errors across platform draws.

The third part of Assumption 2 restricts how error in the platform representations can rotate the true latent platform space. Consider a unit vector $x \in \mathcal{M}$. Then, $\bar{\Gamma}_n x$ describes the average direction of the representation errors, weighted by how strongly those errors align with x. If we project onto the orthogonal complement of the platform space, we then have $P_M^{\perp} \bar{\Gamma}_n x$, which measures the size of the component of the error-induced movement that points outside the latent platform space. Post-multiplying by P_M, to get $P_M^{\perp} \bar{\Gamma}_n P_M$, restricts our attention to vectors in $\mathcal{M}$ as P_M projects a generic vector y into $\mathcal{M}$ before we consider its alignment with the error. The operator norm takes the maximum of this effect over all unit directions, which are first projected into the platform space. Therefore, the third part of Assumption 2 imposes that in the limit, there is no systematic tilt from inside of $\mathcal{M}$ to outside of it from the error.

First, we begin by showing that $\hat{S}_n$ concentrates to its mean. Recall that $S_n^* := \mathbb{E}[\hat{S}_n \mid M_n]$. Then, we have the following result:

Lemma 4. *Under Assumption 2, we have that $\left\| \hat{S}_n - S_n^* \right\| = O_p\left(n^{-1/2}\right)$.*

Proof. Define $Z_j = \tilde{m}_j \tilde{m}_j' - \mathbb{E}[\tilde{m}_j \tilde{m}_j' \mid M_n]$. Conditional on M_n, by Assumption 2 part 2, we have that the Z_j's are independent, symmetric and mean zero matrices. Our goal is to apply a standard matrix concentration result (Vershynin (2018) 5.4.1, Matrix Bernstein). To do so, we show that $\|Z_j\|$ is bounded and give an expression for the corresponding matrix variance term. Note that:

$$\begin{aligned}
\|Z_j\| &= \left\|\tilde{m}_j \tilde{m}_j' - \mathbb{E}[\tilde{m}_j \tilde{m}_j' \mid M_n]\right\| \\
&\leq \left\|\tilde{m}_j \tilde{m}_j'\right\| + \left\|\mathbb{E}[\tilde{m}_j \tilde{m}_j' \mid M_n]\right\| \\
&\leq 1 + \mathbb{E}[\left\|\tilde{m}_j \tilde{m}_j'\right\| \mid M_n] \\
&\leq 1 + 1 = 2
\end{aligned}$$

Where the first step follows by the triangle inequality, the second by the fact $\tilde{m}_j$ is a unit vector and Jensen's inequality, and the third step from $\tilde{m}_j$ being unit.

Then, we can compute:

$$\begin{aligned}
\left\|\sum_{j=1}^{n} \mathbb{E}[Z_j^2 \mid M_n]\right\| &\leq \sum_{j=1}^{n} \left\|\mathbb{E}[Z_j^2 \mid M_n]\right\| \\
&\leq \sum_{j=1}^{n} \mathbb{E}[\left\|Z_j^2\right\| \mid M_n] \\
&\leq \sum_{j=1}^{n} \mathbb{E}[\|Z_j\|^2 \mid M_n] \\
&\leq \sum_{j=1}^{n} \mathbb{E}[4 \mid M_n] = 4n
\end{aligned}$$

Where the first step uses the triangle inequality, the second uses Jensen's inequality, the third the sub-multiplicativity of matrix norms, and the final steps use $\|Z_j\| \leq 2$.

Then, we can apply the Matrix Bernstein inequality for some $C > 0$ (Vershynin (2018) 5.4.1):

$$\begin{aligned}
\mathbb{P}\left(\sqrt{n}\left\|\hat{S}_n - S_n^*\right\| \geq C \mid M_n\right) &= \mathbb{P}\left(\left\|\sum_{j=1}^{n} Z_j\right\| \geq C\sqrt{n} \mid M_n\right) \\
&\leq 2d \exp\left(-\frac{C^2 n/2}{4n + (2C/3)\sqrt{n}}\right)
\end{aligned}$$

Almost surely. For fixed C, as $n \to \infty$, the exponent $\frac{C^2 n/2}{4n+(2C/3)\sqrt{n}} \to \frac{C^2}{8}$. Since d is fixed, we can make the right hand side of this expression arbitrarily small by choosing C sufficiently large. Therefore, we have that $\sqrt{n}\left\|\hat{S}_n - S_n^*\right\| = O_p(1)$, and therefore that $\left\|\hat{S}_n - S_n^*\right\| = O_p(n^{-1/2})$. □

Next, we define $S_n^0 := P_M S_n^* P_M + P_M^\perp S_n^* P_M^\perp$. The matrix S_n^0 is a version of the expected scatter matrix that reproduces the effect of S_n^* on vectors in either $\mathcal{M}$ or $\mathcal{M}^\perp$, but removes the components that map one space to the other.

Lemma 5. *Under Assumption 2:*

(1) *S_n^0 maps $\mathcal{M}$ to itself and $\mathcal{M}^\perp$ to itself.*
(2) $\left\|S_n^* - S_n^0\right\| = o(1)$

Proof. Let $x \in \mathcal{M}$. Then, $P_M x = x$ and $P_M^\perp x = 0$. Therefore, we have that:

$$S_n^0 x = P_M S_n^* P_M x + P_M^\perp S_n^* P_M^\perp x = P_M S_n^* x$$

Because this final expression is pre-multiplied by P_M, the result lies in $\mathcal{M}$, showing $S_n^0 x \in \mathcal{M}$. An analogous result can be shown for $\mathcal{M}^\perp$.

Then to show the second part, begin by pre- and post-multiplying S_n^* by $I_d = P_M + P_M^\perp$, and expand:

$$S_n^* = P_M S_n^* P_M + P_M S_n^* P_M^\perp + P_M^\perp S_n^* P_M + P_M^\perp S_n^* P_M^\perp$$

Subtracting the definition of S_n^0, the two within sub-space terms cancel and we have:

$$S_n^* - S_n^0 = P_M S_n^* P_M^\perp + P_M^\perp S_n^* P_M$$

Therefore, via the triangle inequality, we have that:

$$\left\|S_n^* - S_n^0\right\| \leq \left\|P_M S_n^* P_M^\perp\right\| + \left\|P_M^\perp S_n^* P_M\right\|$$

Noting that because S_n^* is symmetric, we have that $(P_M S_n^* P_M^\perp)' = P_M^\perp S_n^* P_M$, and the operator norm of a matrix is preserved under transpose. Therefore, we have that:

$$\left\|S_n^* - S_n^0\right\| \leq 2\left\|P_M^\perp S_n^* P_M\right\|$$

Finally, to show the result, we re-write S_n^*. Define $\tau_j := \mathbb{E}[\|\eta_j\|^2 \mid M_n]$. Recall that $\tilde{m}_j = a_j m_j + \eta_j$ and that therefore, $\mathbb{E}[a_j^2 \mid M_n] = 1 - \tau_j$ as $a_j^2 + \|\eta_j\|^2 = 1$. We can then compute:

$$\tilde{m}_j \tilde{m}_j' = a_j^2 m_j m_j' + a_j m_j \eta_j' + a_j \eta_j m_j' + \eta_j \eta_j'$$

Furthermore, recall that $\Gamma_j := \mathbb{E}[\eta_j \eta_j' \mid M_n]$. Taking expectations conditional on M_n and applying the first portion of Assumption 2 (the centering condition), we have that:

$$\mathbb{E}[\tilde{m}_j \tilde{m}_j' \mid M_n] = (1 - \tau_j) m_j m_j' + \Gamma_j$$

Averaging, as in S_n^*, we have that:

$$S_n^* = \frac{1}{n} \sum_{j=1}^{n} (1 - \tau_j) m_j m_j' + \bar{\Gamma}_n$$

The first term cannot rotate the latent platform space, as $P_M^{\perp} m_j = 0$ as this computation projects $m_j \in \mathcal{M}$ onto the orthogonal complement of $\mathcal{M}$. Therefore, we have that:

$$P_M^{\perp} \left(\frac{1}{n} \sum_{j=1}^{n} (1 - \tau_j) m_j m_j' \right) P_M = 0$$

Therefore, $P_M^{\perp} S_n^* P_M = P_M^{\perp} \bar{\Gamma}_n P_M$. Recall that by Assumption 2 part 3, we have that $\left\| P_M^{\perp} \bar{\Gamma}_n P_M \right\| = o(1)$. Thus $\left\| S_n^* - S_n^0 \right\| = o(1)$. □

The preceding lemma established that S_n^0 preserves the latent platform space and that deleting the cross components is an asymptotically negligible change to S_n^*. This preservation does not show that $\mathcal{M}$ is the leading eigenspace of S_n^0. To do so, we require an additional condition, that the platform sub-space has sufficient signal relative to its orthogonal complement.

Assumption 3. *Let U_M be a $d \times r$ matrix whose columns form an orthonormal basis for $\mathcal{M}$, and $U_M^{\perp}$ be a $d \times (d - r)$ matrix whose columns form an orthonormal basis for $\mathcal{M}^{\perp}$. Let λ_k be the kth eigenvalue of a matrix. We assume that there exists a constant $\delta > 0$ such that for all sufficiently large n:*

$$\lambda_r(U_M' S_n^* U_M) - \lambda_1((U_M^{\perp})' S_n^* U_M^{\perp}) \geq \delta$$

We can interpret this assumption. Because $U_M' S_n^* U_M$ is a $r \times r$ symmetric matrix, $\lambda_r(U_M' S_n^* U_M)$ is the smallest eigenvalue; therefore, it measures the weakest expected direction in the scatter matrix of the latent platform space. Similarly, $\lambda_1((U_M^\perp)' S_n^* U_M^\perp)$ is the strongest expected scatter of the orthogonal complement of the platform space. Our assumption imposes that the signal in the weakest direction within $\mathcal{M}$ must be stronger than the strongest dimension within $\mathcal{M}^\perp$. This assumption plays two roles: first, it ensures the leading r-dimensional eigenspace of S_n^0 is $\mathcal{M}$, and second, it ensures we have sufficient signal to apply the Davis-Kahan Theorem.

Lemma 6. *Under Assumptions 2 and 3, the leading r-dimensional eigenspace of S_n^0 is exactly $\mathcal{M}$. Furthermore, $\lambda_r(S_n^0) - \lambda_{r+1}(S_n^0) \geq \delta$.*

Proof. Define $Q_M = [U_M \mid U_M^\perp]$. Note that because these matrices are orthonormal bases for complementary spaces, Q_M is orthogonal. We can then use Q_M to perform a change of basis by pre- and post-multiplying, while preserving the eigenstructure. Noting that $P_M = U_M U_M'$ and $P_M^\perp = U_M^\perp (U_M^\perp)'$, the definition of S_n^0 means we have:

$$Q_M' S_n^0 Q_M = \begin{pmatrix} U_M' S_n^* U_M & 0 \\ 0 & (U_M^\perp)' S_n^* U_M^\perp \end{pmatrix}$$

Therefore, S_n^0 is block diagonal with respect to the decomposition of $\mathbb{R}^d$ into $\mathcal{M}$ and $\mathcal{M}^\perp$. Therefore, its eigenvalues are the combined eigenvalues of the two blocks. By Assumption 3, we have that $\lambda_r(U_M' S_n^* U_M) \geq \lambda_1((U_M^\perp)' S_n^* U_M^\perp)$. Therefore, the leading r-dimensional eigenspace of S_n^0 is $\mathcal{M}$. Furthermore, $\lambda_r(S_n^0) = \lambda_r(U_M' S_n^* U_M)$ and $\lambda_{r+1}(S_n^0) = \lambda_1((U_M^\perp)' S_n^* U_M^\perp)$, so we have $\lambda_r(S_n^0) - \lambda_{r+1}(S_n^0) \geq \delta > 0$. □

Finally, we combine our results, in conjunction with the Davis-Kahan Theorem, in order to demonstrate that projection onto the first r left singular vectors of $\tilde{M}_n$ converges to projecting onto $\mathcal{M}$. We state the version of the Davis-Kahan Theorem that we use, which is due to Yu et al. (2015). This version is convenient for our purposes because it requires an eigengap only in the reference matrix[27], which in our case is S_n^0, and not in the perturbed matrix $\hat{S}_n$. We state a modified version of the theorem from Yu et al. (2015), adapted to our setting.

Theorem 7 (Davis-Kahan, Yu et al. (2015) Theorem 2)**.** *Let A and $B = A + E$ be symmetric $d \times d$ matrices, and suppose that $\lambda_r(A) - \lambda_{r+1}(A) \geq \delta > 0$. Let P_A denote the projection onto the leading*

[27]Note that the theorem makes no claim about B having an eigengap of its own. If $\lambda_r(B) = \lambda_{r+1}(B)$, then the leading r-dimensional eigenspace of B is not unique. However, the theorem states that every such choice is close to the leading eigenspace of A. We therefore do not need to establish that $\hat{S}_n$ has an eigengap, which lets us avoid an additional perturbation argument.

r-dimensional eigenspace of A, and let P_B denote the projection onto the span of any r orthonormal eigenvectors of B corresponding to its r largest eigenvalues. Then:

$$\|P_B - P_A\| \leq \frac{2\sqrt{r}\,\|E\|}{\delta}$$

The Davis-Kahan Theorem is typically stated in terms of $\|\sin\Theta\|$, where Θ is the diagonal matrix of principal angles between the two eigenspaces. Because the operator norm of the difference of two orthogonal projections onto subspaces of equal dimension is exactly the sine of the largest principal angle between them (Chen et al. (2021) Lemma 2.5), the two formulations are equivalent. We focus on the projection version, as this object is what we need to compute political content and slant.

We now combine the preceding results to show that projection onto the leading r-dimensional eigenspace of $\hat{S}_n$ converges to projection onto the latent platform space $\mathcal{M}$.

Proposition 8. *Under Assumptions 2 and 3, we have that:*

$$\left\|\hat{P}_M - P_M\right\| = o_p(1).$$

Proof. We begin by bounding the distance between $\hat{S}_n$ and S_n^0. Adding and subtracting S_n^*:

$$\hat{S}_n - S_n^0 = \underbrace{\hat{S}_n - S_n^*}_{\text{Random fluctuation}} + \underbrace{S_n^* - S_n^0}_{\text{Deterministic Rotation}}$$

By the triangle inequality, we have that:

$$\left\|\hat{S}_n - S_n^0\right\| \leq \left\|\hat{S}_n - S_n^*\right\| + \left\|S_n^* - S_n^0\right\|$$

Then, we can apply Lemmas 4 and 5:

$$\begin{aligned}\left\|\hat{S}_n - S_n^0\right\| &= O_p(n^{-1/2}) + o(1)\\ &= o_p(1)\end{aligned}$$

We now apply Theorem 7. By Lemma 6, the leading r-dimensional eigenspace of S_n^0 is exactly $\mathcal{M}$. Therefore, the projection onto the leading r eigenvectors of S_n^0 is P_M. Lemma 6 also establishes that:

$$\lambda_r(S_n^0) - \lambda_{r+1}(S_n^0) \geq \delta.$$

We can then apply Theorem 7 with $A = S_n^0$, $B = \hat{S}_n$, and $E = \hat{S}_n - S_n^0$. By definition, the projection onto the leading r-dimensional eigenspace of $\hat{S}_n$ is $\hat{P}_M$. Therefore:

$$\left\| \hat{P}_M - P_M \right\| \leq \frac{2\sqrt{r}}{\delta} \left\| \hat{S}_n - S_n^0 \right\|.$$

Because r and δ are fixed and $\left\| \hat{S}_n - S_n^0 \right\| = o_p(1)$, it follows that:

$$\left\| \hat{P}_M - P_M \right\| = o_p(1).$$

Finally, by Lemma 3, the leading r eigenvectors of $\hat{S}_n = \frac{1}{n}\tilde{M}_n\tilde{M}_n'$ are exactly the leading r left singular vectors of $\tilde{M}_n$. Therefore, $\hat{P}_M$ is exactly the projection matrix produced by applying our empirical SVD procedure to the observed platform representation matrix. □

We can then apply Proposition 8 separately to the combined platform matrix $X_n = [R_n \mid D_n]$, the Republican platform matrix R_n, and the Democratic platform matrix D_n. Let P_X, P_R, and P_D denote the corresponding latent projection matrices, and let $\hat{P}_X$, $\hat{P}_R$, and $\hat{P}_D$ denote the projections recovered from the observed platform representations. Under Assumptions 2 and 3 for each of these matrices, Proposition 8 implies:

$$\left\| \hat{P}_X - P_X \right\| = o_p(1), \qquad \left\| \hat{P}_R - P_R \right\| = o_p(1), \qquad \left\| \hat{P}_D - P_D \right\| = o_p(1).$$

We can therefore establish consistency of political content and slant constructed using the observed platform representations. In the empirical construction, we first project the course representation onto the combined political subspace and then compute its alignment with each party subspace. Thus, we define estimators for political content and slant as:

$$\hat{p}(c \mid \tilde{R}, \tilde{D}) = \|\hat{P}_X c\|^2,$$
$$\hat{s}(c \mid \tilde{R}, \tilde{D}) = \|\hat{P}_R \hat{P}_X c\|^2 - \|\hat{P}_D \hat{P}_X c\|^2.$$

Corollary 1. *For any unit course representation c, political content and slant computed using the observed platform representations converge in probability to political content and slant computed using the latent platform representations. In particular:*

$$\hat{p}(c \mid \tilde{R}, \tilde{D}) - p(c \mid R, D) = o_p(1), \qquad \hat{s}(c \mid \tilde{R}, \tilde{D}) - s(c \mid R, D) = o_p(1).$$

Moreover, both results hold uniformly over all unit course representation vectors c.

Proof. By the projection identities above, political content can be written as $p(c \mid R, D) = c' P_X c$, with the analogous expression $\hat{p}(c \mid \tilde{R}, \tilde{D}) = c' \hat{P}_X c$ using the observed platforms. Therefore, Proposition 8 implies

$$\left| \hat{p}(c \mid \tilde{R}, \tilde{D}) - p(c \mid R, D) \right| \leq \left\| \hat{P}_X - P_X \right\| = o_p(1).$$

For slant, our empirical construction gives

$$\hat{s}(c \mid \tilde{R}, \tilde{D}) = c'(\hat{P}_X \hat{P}_R \hat{P}_X - \hat{P}_X \hat{P}_D \hat{P}_X)c.$$

At the latent level, the Republican and Democratic platform spaces are contained in the combined platform space, so $P_X P_R P_X = P_R$ and $P_X P_D P_X = P_D$. For the Republican component, adding and subtracting intermediate terms gives

$$\hat{P}_X \hat{P}_R \hat{P}_X - P_R = (\hat{P}_X - P_X)\hat{P}_R \hat{P}_X + P_X(\hat{P}_R - P_R)\hat{P}_X + P_X P_R(\hat{P}_X - P_X),$$

Then, via the triangle inequality, sub-multiplicativity of norms, and the fact that projection matrices have operator norm at most one, we have that:

$$\left\| \hat{P}_X \hat{P}_R \hat{P}_X - P_R \right\| \leq 2 \left\| \hat{P}_X - P_X \right\| + \left\| \hat{P}_R - P_R \right\|.$$

Analogously, we can compute:

$$\left\| \hat{P}_X \hat{P}_D \hat{P}_X - P_D \right\| \leq 2 \left\| \hat{P}_X - P_X \right\| + \left\| \hat{P}_D - P_D \right\|.$$

Then via the triangle inequality, we have that:

$$\left| \hat{s}(c \mid \tilde{R}, \tilde{D}) - s(c \mid R, D) \right| \leq 4 \left\| \hat{P}_X - P_X \right\| + \left\| \hat{P}_R - P_R \right\| + \left\| \hat{P}_D - P_D \right\| = o_p(1),$$

where the final result follows from Proposition 8 applied separately to the combined, Republican, and Democratic platform matrices. Because neither bound depends on c, both convergence results hold uniformly over all unit course representation vectors. $\square$

Finally, we can combine the results on measurement error in the course representations and platform representations. Consider political content computed using both the observed course representation and the observed platform representations. Adding and subtracting political content computed using the observed course representation and latent platform space, we have:

$$\hat{p}(\tilde{c}_i \mid \tilde{R}, \tilde{D}) - p(c_i \mid R, D) = \left[\hat{p}(\tilde{c}_i \mid \tilde{R}, \tilde{D}) - p(\tilde{c}_i \mid R, D)\right] + \left[p(\tilde{c}_i \mid R, D) - p(c_i \mid R, D)\right].$$

By Corollary 1, the first term is $o_p(1)$ uniformly over all unit course representation vectors. Because $\tilde{c}_i$ is unit length, this result also applies when evaluating the measure at the observed course representation $\tilde{c}_i$. By Proposition 2, the second term can be written as $\kappa_p + u_i^p$, where $\mathbb{E}[u_i^p \mid c_i, \Omega_i] = 0$. Therefore:

$$\hat{p}(\tilde{c}_i \mid \tilde{R}, \tilde{D}) = p(c_i \mid R, D) + \kappa_p + u_i^p + o_p(1).$$

An analogous argument applies to slant. Corollary 1 implies that the difference between slant computed using the observed and latent platform spaces is $o_p(1)$ uniformly over unit course representation vectors, while Proposition 2 implies that measurement error from the course representation can be written as $\kappa_s + u_i^s$, where $\mathbb{E}[u_i^s \mid c_i, \Omega_i] = 0$. Therefore:

$$\hat{s}(\tilde{c}_i \mid \tilde{R}, \tilde{D}) = s(c_i \mid R, D) + \kappa_s + u_i^s + o_p(1).$$

Thus, when both the course and platform representations are observed with error, the additional error from estimating the platform spaces is asymptotically negligible, while the remaining course-level measurement error takes the constant plus conditionally mean-zero form characterized in Proposition 2.

A.2.3. *Comparison to Supervised Learning.* Next, we consider a stylized comparison to more standard supervised learning methods used to assess partisan slant in text (Gentzkow and Shapiro, 2010; Martin and Yurukoglu, 2017). This style of measuring partisan slant of text proceeds in two steps. First, the researcher trains a model on text with known partisan labels; as a concrete example, both Gentzkow and Shapiro (2010) and Martin and Yurukoglu (2017) use speeches from the Congressional Record, with labels coming from the party affiliation of the speaker. Second, the researcher takes the trained model and predicts the partisan slant of their out of sample text. We call this the supervised slant, to differentiate between this method common in the literature and our method.

For the sake of direct comparison between methods, we consider a training set based on the same set of labeled Republican and Democratic party platform vectors. We consider the dataset $\{y_i, b_i\}$ where y_i is a label for embedding vector i, denoting if i is a Republican or Democratic party platform vector, and b_i is the representation of the platform section. Note that by construction, $b_i \in \text{span}(X)$, the space spanned by all of the platform vectors.[28]

We focus on the class of supervised learning problems characterized by a linear index. Formally, for a representation c, the loss function for the supervised learning problem first computes $\alpha + \beta' c$ for parameters α and β, before plugging this value into the loss function.

$$L_n(\alpha, \beta; \{y_i, b_i\}) = Q(\{y_i\}, \{\alpha + \beta' b_i\})$$

For some function Q. Many cases impose that Q is additively separable across data points; specific examples include squared error and logistic loss. The researcher then estimates $\hat{\beta}$ by minimizing the loss function. Out of sample, the researcher then computes some function $g(\hat{\alpha} + \hat{\beta}' c)$ as the supervised slant of c.

Proposition 9. *Consider a set of parameters $(\hat{\alpha}, \hat{\beta})$ that minimizes $L_n(\alpha, \beta; \{y_i, b_i\})$. Consider an out of sample vector c with amount of political content not equal to one, i.e. $c \notin \text{span}(X)$. Then with equal in sample loss, the supervised slant can take any value in the range of g.*

Proof. Consider fitted parameters from minimizing the loss function $(\hat{\alpha}, \hat{\beta})$. First, note that for any $z \in X^{\perp}$, the set orthogonal to X, $(\hat{\alpha}, \hat{\beta} + z)$ achieves equivalent loss. This fact follows as the linear indices of each data point are identical. For a data point b_i:

$$\hat{\alpha} + \hat{\beta}' b_i + \underbrace{z' b_i}_{0 \text{ by } z \in X^{\perp}} = \hat{\alpha} + \hat{\beta}' b_i$$

Therefore, we can consider the set of loss-minimizing parameters $(\hat{\alpha}, \hat{\beta} + z)$.

Consider prediction out of sample for a representation vector $c \notin X$. Let $c_{\perp} := (I - P_X)c$, the part of c not in X. As c is not in X, we have $c_{\perp} \neq 0$. For any $t \in \mathbb{R}$, set $z_t = t c_{\perp}$; note that because $c_{\perp}$ is in the set of vectors orthogonal to X, so is z_t. Therefore, $(\hat{\alpha}, \hat{\beta} + z_t)$ achieves the same training loss. Then, the supervised slant of c is:

$$\begin{aligned}
\hat{\alpha} + (\hat{\beta} + z_t)' c &= \hat{\alpha} + \hat{\beta}' c + t c_{\perp}' c \\
&= \hat{\alpha} + \hat{\beta}' c + t c_{\perp}' [P_X c + (I - P_X) c] \\
&= \hat{\alpha} + \hat{\beta}' c + t c_{\perp}' c_{\perp} + t c_{\perp}' P_X c
\end{aligned}$$

[28] Numerically, this may not be exactly the case once we have applied SVD to the platforms. However, we abstract from this numerical issue for the sake of this comparison.

$$= \hat{\alpha} + \hat{\beta}' c + t\|c_\perp\|^2.$$

Where we note that $c_\perp' P_X c = 0$ as $P_X c$ is in X by construction. Because $\|c_\perp\|^2 > 0$, this expression ranges over all of $\mathbb{R}$ as t ranges over $\mathbb{R}$. Therefore, the supervised slant can be made arbitrary within its range for observationally equivalent loss, and thus is not identified. □

This non-identification result points to an issue with the use of supervised learning when the amount of political content is heterogeneous, in the specific case of loss functions with a linear index. Any form of regularization may select a particular member of the equivalence class of loss functions, but this selection supplies further information beyond how well the classifier can predict the document labels.

A.3. **Measurement Implementation and Data Cleaning.** Given the results in Section A.2, we outline the following procedure to construct our measurements. We elaborate on steps in the following subsections.

(1) Embed sections of the Republican and Democratic platforms.
(2) Horizontally concatenate the vectors to form $\tilde{X}$, $\tilde{R}$, and $\tilde{D}$.
(3) Using SVD, rank reduce each matrix to obtain $\hat{X}$, $\hat{R}$, and $\hat{D}$.
(4) Embed the course description vectors after removing boilerplate text.
(5) Compute political content and slant via the estimators defined in Section A.2; valence is computed as slant divided by political content.

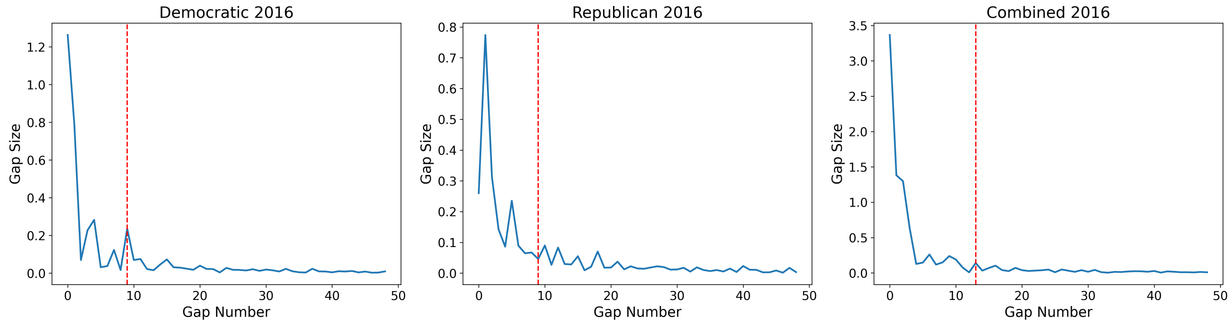
A.3.1. *SVD and Rank Selection.* The SVD procedure in Section A.2 requires the rank of the subspace. We choose the rank to correspond to explaining 65% of the variance in the platform embedding vectors, based on the singular value gaps shown in Figure A.4. Given the singular value separation used in the proofs (Assumption 3), we intentionally do not explain all variance.

We show the correlation of political content and slant using an 80% threshold versus the baseline 65% in Figure A.5.

A.3.2. *Course Description Cleaning.* To strip out boilerplate text in the course descriptions, we proceed in three steps.

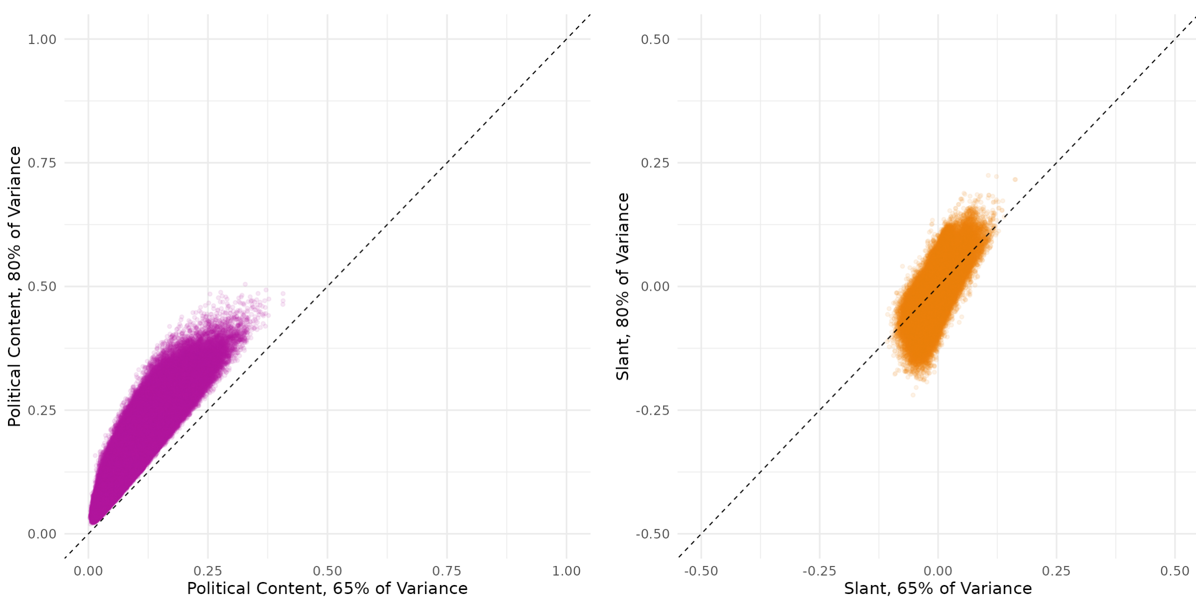
(1) We divide course descriptions into sentences using the `en core web sm` model from `spacy`

FIGURE A.4. Singular value gaps and rank selection for the 2016 platforms



Notes: The figure reports singular values from the SVD of the 2016 Democratic platform, Republican platform, and the horizontally concatenated Democratic-Republican platform matrix used to construct the political subspace. The vertical line marks the rank at which 65% of the variance in the platform embeddings is explained, which is the rank used in the main specification.

FIGURE A.5. Sensitivity of political content and slant to the SVD variance threshold



(A) Political content

(B) Slant

Notes: Each panel compares course-level political content (left) or slant (right) computed with a 65% variance threshold (the baseline) against an 80% threshold in the SVD step used to construct the political subspace.

(2) All sentences containing a set of pre-defined phrases are removed. Pre-defined phrases include terms like "prerequisite," "instructor's permission," and "grade of c-." The full list will eventually be made available in a replication package.

(3) All sentences that appear verbatim in greater than 1% of course descriptions within an institution are also removed.

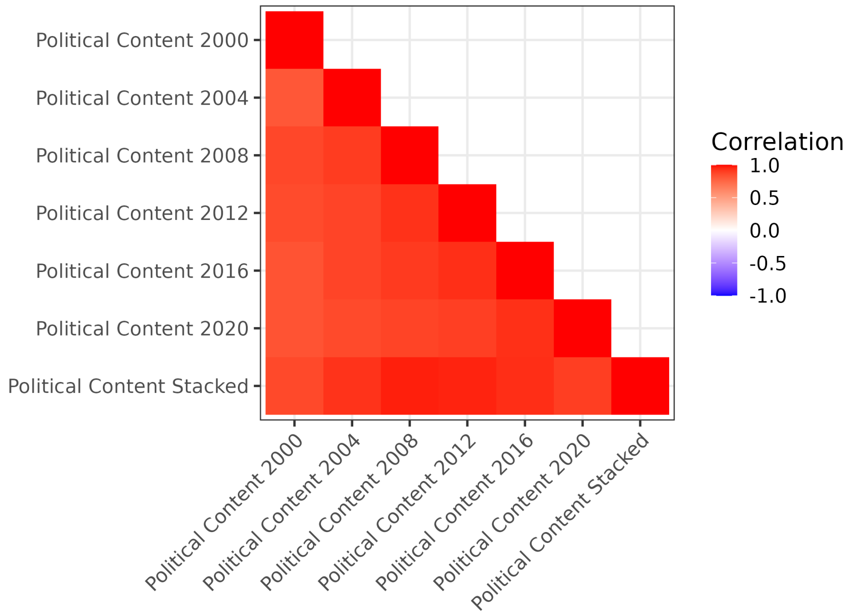
A.4. **Contemporaneous Platforms.** Here, we consider specifications of the main descriptive facts with alternative platforms.

Our "rolling specification," which we introduce in Section 4.4, constructs our political characteristic measures by comparing course description text to the prevailing party platforms rather than the fixed 2016 platform. Specifically, each platform p covers academic years $\{p-2,\ p-1,\ p,\ p+1, p+2\}$, which generates some overlap. For example, the 2016 platforms apply to courses offered in academic years 2014 through 2018, and the 2020 platforms to courses offered in 2018 through 2022. For courses in boundary years between platforms (e.g., 2018 for the transition from the 2016 to 2020 platforms), we include two observations: one based on comparisons to the previous platform and one based on the new platform, which allows us to identify platform-by-institution-by-field fixed effects. Because different periods of time are now measured against different reference platforms, we allow institution-by-field fixed effects to vary by platform ($\kappa_{ifp(t)}$), so that demeaning occurs within each platform.

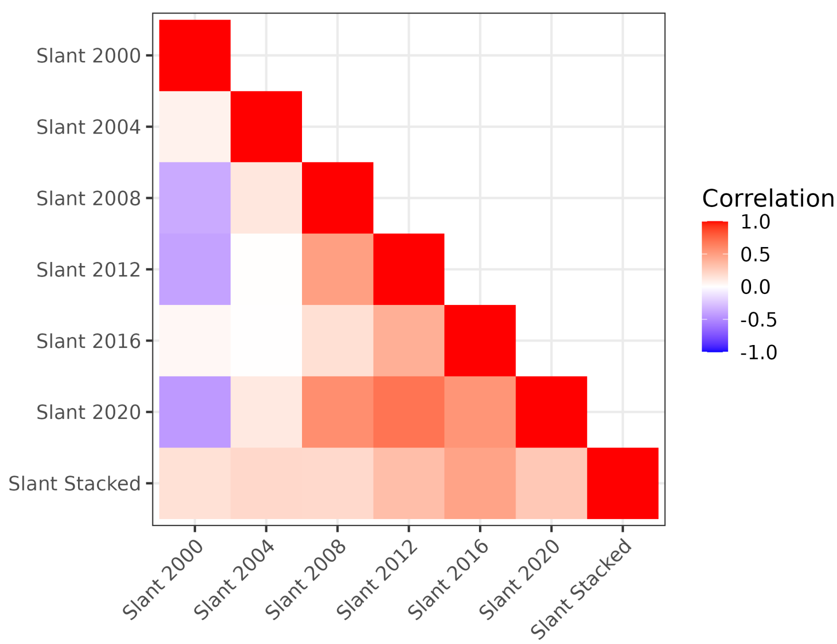
Figure A.6 shows that the measures are generally highly correlated, no matter which platform we use as a comparison. Rolling refers to comparison to contemporaneous platforms, while stacked refers to using the text from all party platforms from 2000 to 2024 as the comparison text. Political content is highly correlated across all platform baselines. For slant, most variants are highly positively correlated. The one exception to the positive correlation is the set of 2000 platforms for slant, which is negatively correlated with most other platforms.

Then, we run the same regressions as in Section 4, but use contemporaneous political platforms. As an additional control, we include an indicator for which platform is used for comparisons. Results are shown in Figure A.7. We interact the platform fixed effects with the non-over time controls. Vertical dotted lines show changes in which platform is used as comparison.

FIGURE A.6. Correlation of political content and slant across benchmark platforms



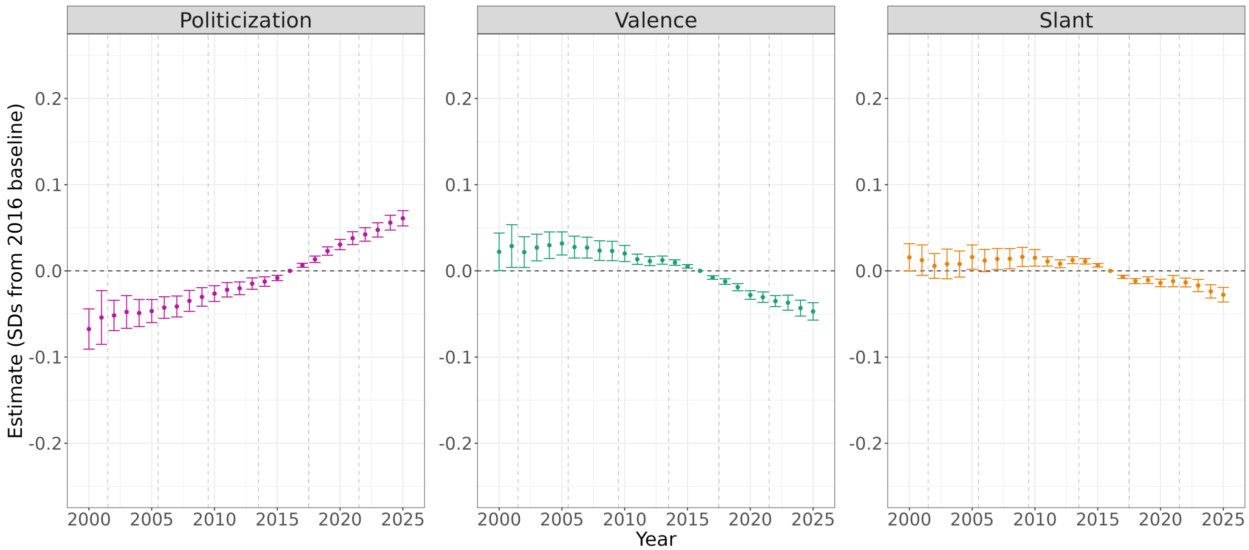
(A) Political Content Correlation



(B) Slant Correlation

Notes: Each panel reports pairwise correlations of course political content (top) or slant (bottom) computed under alternative benchmark documents. Year labels refer to the party platforms of that year. "Stacked" uses the combined text from all platforms as the benchmark.

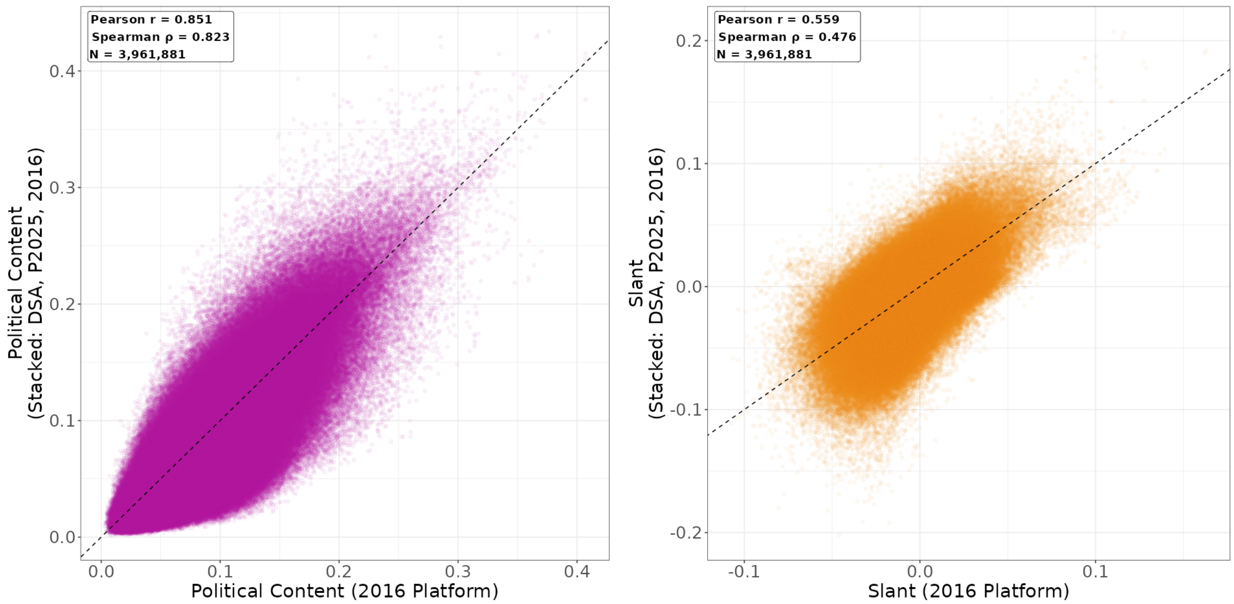
FIGURE A.7. Trends in political content, valence, and slant, rolling (contemporaneous) platforms



Notes: The figure reproduces the rolling-platform trend specification in Figure 6, adding valence as the middle panel. Political content, valence, and slant are constructed using the party platforms assigned to each course's time period rather than the fixed 2016 benchmarks. Points are coefficients on academic-year indicators from regressions controlling for institution-by-field-by-platform fixed effects. Measures are demeaned within platform and standardized using the cross-course standard-deviation in the corresponding platform year. Boundary years between platforms enter once under each adjacent benchmark, allowing the platform-specific fixed effects to be identified. Standard errors are clustered at the institution level; vertical bars are 95% confidence intervals.

A.5. **Alternate Platforms.** As an additional robustness check, we recompute our measures using broader definitions of liberal and conservative benchmarks. Specifically, we augment our party platform benchmarks with similar political documents from groups considered to be more extreme. For the liberal benchmark, we horizontally concatenate the embedding vectors for the 2016 Democratic platform with embedding vectors based on sections of the 2021 Democratic Socialists of America platform, to capture political characteristics to the left of the Democratic Party. Similarly, we concatenate sections from the introduction of Project 2025's "Mandate for Leadership" from 2023 with the 2016 Republican Party platform. We repeat the same SVD process as in the main specification using just the Democratic and Republican platforms, but with the combined benchmarks. Again based on the singular value gaps, we choose a rank of 9 for both combined benchmarks (Republican and Project 2025, and Democratic and DSA respectively). We choose a rank of 12 for the combined political space. The Pearson correlation between values for political content is 0.84, while for slant

FIGURE A.8. Correlation of base measures to measures constructed with augmented benchmark (2016 platforms, Project 2025, DSA platform)

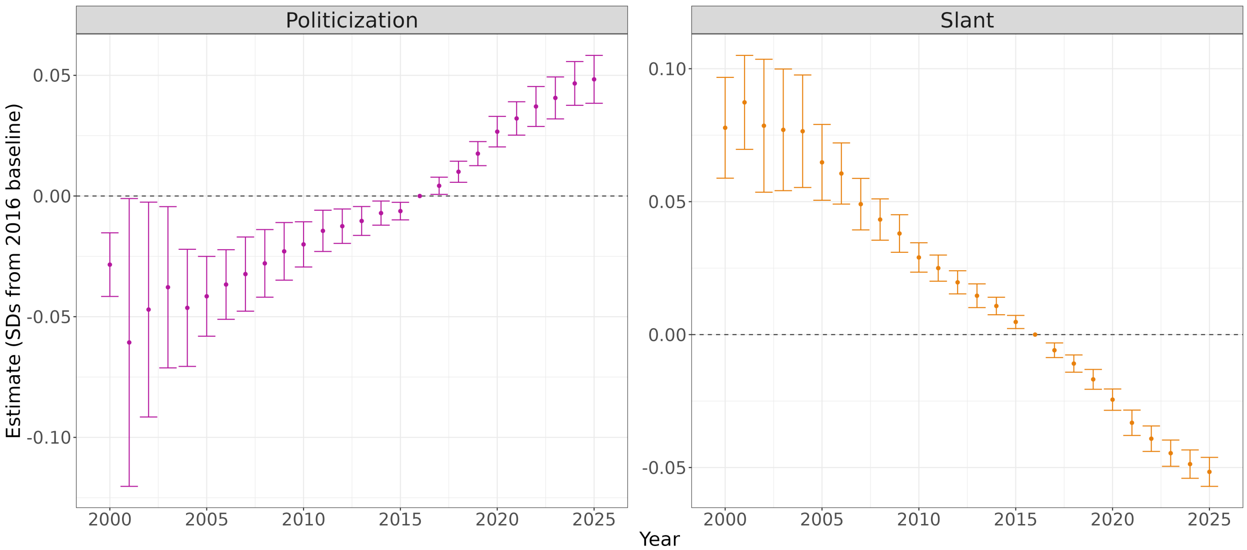


(A) Political content (B) Slant

Notes: Each panel compares course-level measures under the baseline 2016 party-platform benchmarks (horizontal axis) with measures under augmented benchmarks (vertical axis). The augmented liberal benchmark combines the 2016 Democratic platform with the Democratic Socialists of America platform; the augmented conservative benchmark combines the 2016 Republican platform with the introduction to Project 2025's "Mandate for Leadership." Panel (a) reports political content and Panel (b) slant. Pearson correlations between the baseline and augmented measures are 0.84 for political content and 0.56 for slant.

it is 0.56, and the scatter plots are shown in Figure A.8. We show the analogous version of Figure 4 in Figure A.9.

Figure A.9. Trends in political content and slant, augmented benchmark (2016 platforms, Project 2025, DSA platform)



Notes: The figure reproduces the specification in Figure 4 using augmented partisan benchmarks. The liberal benchmark combines the 2016 Democratic platform with the Democratic Socialists of America platform, while the conservative benchmark combines the 2016 Republican platform with the introduction to Project 2025's "Mandate for Leadership." Points are coefficients on academic-year indicators from regressions of political content (left) and slant (right) controlling for institution-by-field fixed effects. Courses offered in 2016-17 are the omitted baseline, and outcomes are expressed in cross-course standard-deviation units. Standard errors are clustered at the institution level; vertical bars are 95% confidence intervals.

Appendix B. Descriptives Supplement

This appendix contains supplementary material for Section 4, focusing on descriptive patterns in our measures.

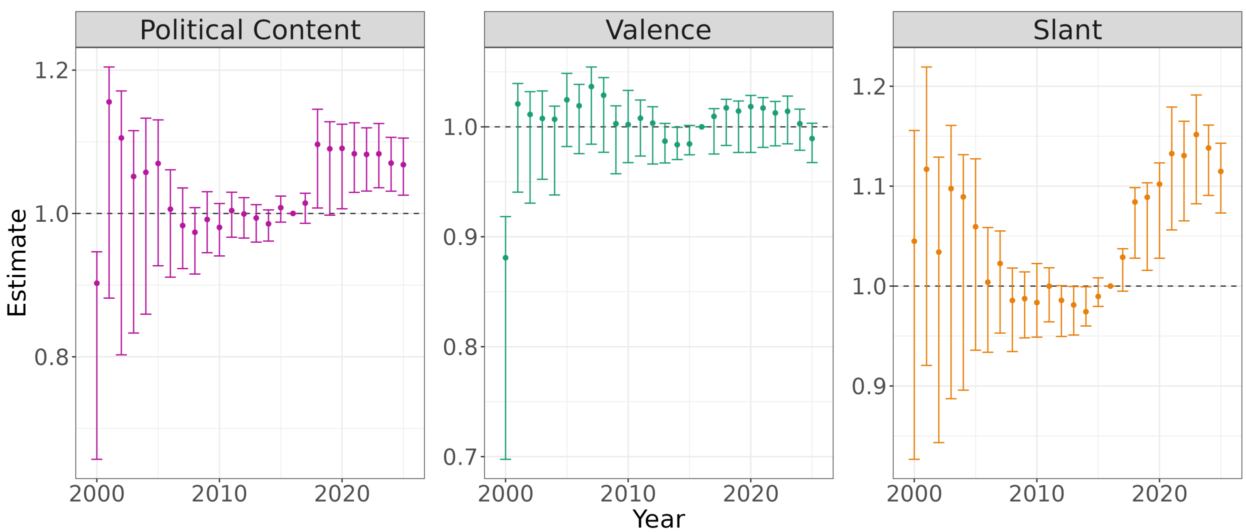
B.1. **Supplementary Exhibits.**

Table B.1. Fields to field group mapping

Field group	Fields
Arts & Language	English, Language, Arts
Business	Business, Hospitality, Human Resources, Consumer Science, Economics
STEM	Agriculture, Biology, Environmental Studies, Engineering, Computer Science, Math, Physics, Chemistry, Stats/Data Science, Science - Other
Social Sciences	Anthropology, Sociology, Psychology, Political Science, Social Science, Criminal Justice, International Studies, Security Studies, Public Policy, Social Work
Humanities	Philosophy, History, Religion, American Studies, Ethnic/Cultural Studies, Community Studies, Humanities

Notes: The table lists the fields grouped into each field group used throughout the paper. Fields not listed (primarily professional programs) are not assigned to a field group and are excluded only from analyses that require these broader categories; they remain in analyses using the full course sample.

Figure B.1. Trend in variance of political content and slant



Notes: For each academic year and content measure, the figure computes the variance across courses within each institution-by-field cell and then averages these variances across institution-field cells. Values are normalized by the corresponding average within-cell variance in 2016-17, so a value of one denotes 2016-17 dispersion and a value above one denotes greater within-institution-field dispersion. Confidence intervals are obtained from 100 block bootstrap replications that resample institutions; vertical bars report 95% confidence intervals.

Table B.2. Average political content and slant by field

		Political Content		Slant	
Field / department	Courses		Norm.		Norm.
Social Sciences	113,752	0.0823	+0.53	-0.0061	-0.53
Political Science	17,625	0.1198	+1.61	-0.0003	-0.04
Criminal Justice	11,464	0.0906	+0.77	-0.0036	-0.32
Sociology	14,913	0.0788	+0.43	-0.0088	-0.75
Anthropology	10,677	0.0656	+0.04	-0.0127	-1.08
Psychology	30,624	0.0594	-0.14	-0.0022	-0.20
Humanities	66,864	0.0750	+0.31	-0.0014	-0.13
Ethnic/Cultural Studies	15,138	0.0893	+0.73	-0.0134	-1.14
History	23,552	0.0791	+0.43	-0.0036	-0.31
Religion	10,589	0.0646	+0.01	0.0070	+0.57
Humanities	5,474	0.0615	-0.07	0.0015	+0.11
Philosophy	10,843	0.0607	-0.10	0.0118	+0.98
Business	79,891	0.0715	+0.21	0.0015	+0.11
Economics	12,598	0.0957	+0.91	-0.0003	-0.04
Consumer Science	2,346	0.0712	+0.21	0.0000	-0.01
Human Resources	1,295	0.0709	+0.20	-0.0094	-0.80
Business	59,687	0.0676	+0.10	0.0024	+0.19
Hospitality	3,965	0.0539	-0.30	-0.0031	-0.27
STEM	226,002	0.0536	-0.30	0.0018	+0.14
Engineering	45,585	0.0570	-0.20	0.0027	+0.22
Computer Science	43,868	0.0539	-0.29	0.0057	+0.47
Biology	34,629	0.0504	-0.40	-0.0007	-0.07
Chemistry	16,264	0.0401	-0.69	0.0019	+0.15
Math	24,300	0.0383	-0.74	0.0052	+0.43
Arts & Language	192,376	0.0443	-0.57	0.0017	+0.13
English	35,112	0.0503	-0.40	0.0013	+0.10
Arts	113,675	0.0440	-0.58	0.0033	+0.27
Language	43,589	0.0401	-0.69	-0.0024	-0.21

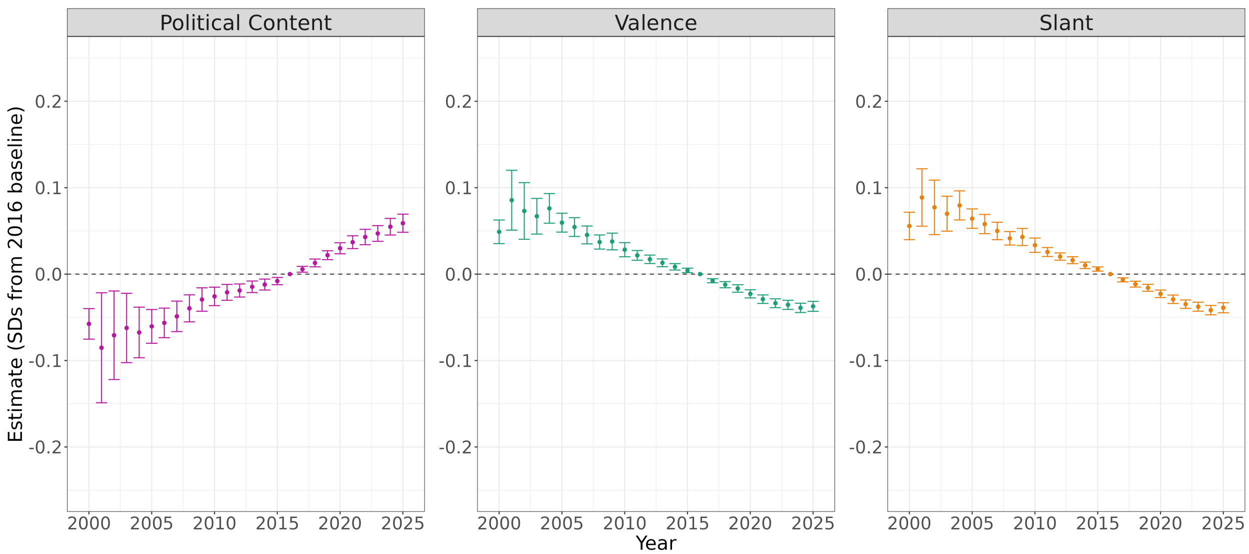
Notes: The table reports mean political content and slant by field using courses offered in 2016-17. Within each field group, the table includes the five fields with the largest number of courses, or all fields when the group contains fewer than five. Fields are ordered by mean political content. Raw measures are reported alongside standardized values expressed relative to the cross-course 2016-17 distribution.

Table B.3. Classification of states based on voting history

Classification	States
Red	AL, AK, AR, FL, ID, IN, IA, KS, KY, LA, MS, MO, MT, NE, NC, ND, OH, OK, SC, SD, TN, TX, UT, WV, WY
Blue	CA, CO, CT, DE, DC, HI, IL, ME, MD, MA, MN, NH, NJ, NM, NY, OR, RI, VT, VA, WA

Notes: The table reports the state classification used in analyses of heterogeneity by state politics. A state is classified as Blue if the Democratic presidential candidate won the state in each of the 2016, 2020, and 2024 elections and Red if the Republican candidate won in all three elections. States that switched parties across these elections are not classified and are excluded only from analyses using the Red/Blue classification.

Figure B.2. Trends in political content, valence, and slant



Notes: Points are coefficients from a regression of political content (left panel), valence (middle panel), and slant (right panel) on year indicators, controlling for institution-by-field fixed effects. Courses offered in 2016-17 are the omitted baseline, so each coefficient is a deviation from the 2016 within-institution-by-field mean, measured in 2016-17 course standard deviation units. Political content measures the extent to which a course engages political language; valence is the relative proximity to each partisan pole within the political subspace without accounting for whether the course is political in the first place; slant measures the partisan orientation of that engagement, with negative values indicating liberal alignment. Standard errors are clustered at the institution level; vertical bars are 95% confidence intervals.

TABLE B.4. Fitted levels and 25-year trends in political content and slant, with standard errors

	Political Content			Slant		
	Means		Δ (25 yr)	Means		Δ (25 yr)
	2000-01	2025-26	Slope	2000-01	2025-26	Slope
Overall	—	0.15	0.15	—	-0.13	-0.13
			(0.013)			(0.007)
Field group						
STEM*	—	0.18	0.18	—	-0.07	-0.07
			(0.014)			(0.007)
Business	0.59	0.64	0.05	-0.00	-0.11	-0.11
			(0.017)			(0.010)
Arts & Language	-0.27	-0.09	0.18	0.01	-0.10	-0.11
			(0.013)			(0.009)
Humanities	0.59	0.82	0.22	-0.13	-0.43	-0.30
			(0.024)			(0.023)
Social Sciences	0.87	0.99	0.12	-0.62	-0.78	-0.16
			(0.018)			(0.013)
Selectivity						
Highly Selective*	—	0.28	0.28	—	-0.21	-0.21
			(0.026)			(0.019)
Mid Selective	-0.04	0.11	0.15	0.05	-0.09	-0.14
			(0.012)			(0.009)
Least Selective	-0.03	0.07	0.10	0.09	-0.01	-0.09
			(0.021)			(0.008)
Institution control						
Public*	—	0.13	0.13	—	-0.11	-0.11
			(0.014)			(0.007)
Private	0.07	0.27	0.21	0.03	-0.15	-0.18
			(0.024)			(0.013)
State politics						
Red*	—	0.15	0.15	—	-0.11	-0.11
			(0.015)			(0.010)
Blue	0.06	0.22	0.15	-0.05	-0.21	-0.16
			(0.026)			(0.010)

Notes: This table reproduces Table 4 with standard errors. For each subgroup and outcome, estimates come from a regression on a linear academic year trend with institution-by-field fixed effects. The 2000-01 and 2025-26 columns report fitted values evaluated at those years, and the 25-year trend equals the estimated annual slope multiplied by 25. All outcomes are expressed in 2016-17 cross-course standard deviation units. Within each panel, levels are reported relative to the 2000-01 level of the first category (indicated with an *). Standard errors are clustered at the institution level; all slope estimates are significant at the 1% level.

Table B.5. Mapping of quantiles in 2005-10 to quantiles in 2020-25

	Political Content		Slant	
p	2020-2025 share below (%)	Δ tail mass	2020-2025 share below (%)	Δ tail mass
1	0.55*** (0.07)	-44.9%	1.49*** (0.05)	+48.6%
5	3.34*** (0.20)	-33.1%	6.31*** (0.17)	+26.2%
25	20.49*** (0.49)	-18.0%	27.77*** (0.44)	+11.1%
50	45.76*** (0.54)	-8.5%	52.40*** (0.57)	+4.8%
75	72.88*** (0.44)	+8.5%	76.63*** (0.44)	-6.5%
95	94.48*** (0.19)	+10.4%	95.60*** (0.14)	-11.9%
99	98.93 (0.07)	+7.4%	99.19*** (0.04)	-19.5%

Notes: The second and fourth columns show the 2020-2025 mass below the 2005-2010 percentile cutoff dictated by percentile p in the first column. Two stars represent a change significant at the $p < 0.05$ level; three stars, at the $p < 0.01$ level. The third and fifth columns show the change in mass *relative* to the 2005-2010 baseline. Tail mass changes below the median represent changes in the lower tail; changes above the median represent changes in the upper tail. Table B.6 produces the same results using the balanced panel of 192 schools.

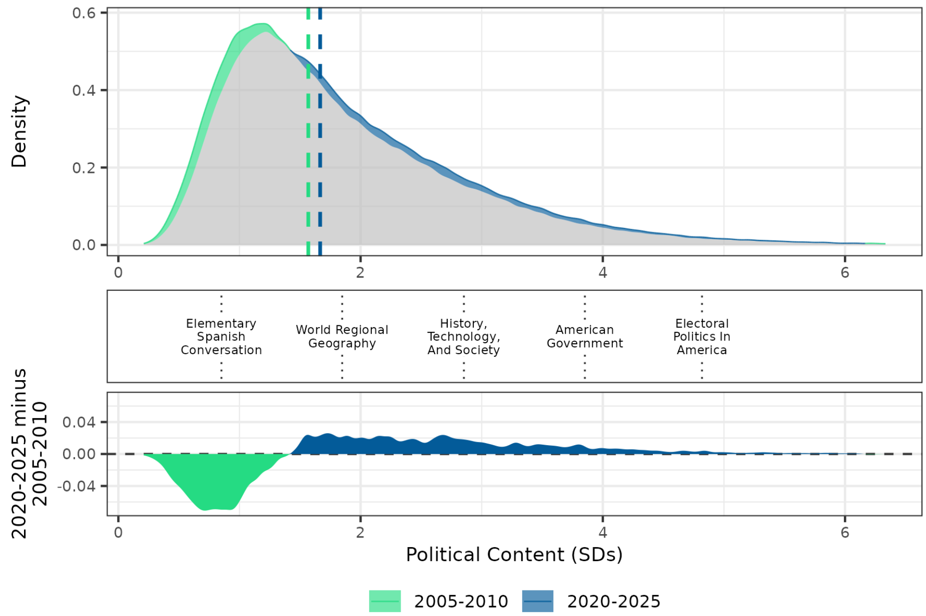
Table B.6. Mapping of quantiles in 2005-2010 to quantiles in 2020-2025, balanced panel

	Political Content		Slant	
p	2020-2025 share below (%)	Δ tail mass	2020-2025 share below (%)	Δ tail mass
1	0.63*** (0.05)	-36.6%	1.53*** (0.05)	+53.3%
5	3.57*** (0.16)	-28.7%	6.49*** (0.12)	+29.7%
25	20.85*** (0.45)	-16.6%	28.19*** (0.30)	+12.7%
50	45.52*** (0.52)	-9.0%	52.71*** (0.37)	+5.4%
75	72.11*** (0.43)	+11.6%	76.62*** (0.30)	-6.5%
95	94.19*** (0.16)	+16.2%	95.49*** (0.10)	-9.8%
99	98.85** (0.06)	+15.2%	99.13*** (0.03)	-13.0%

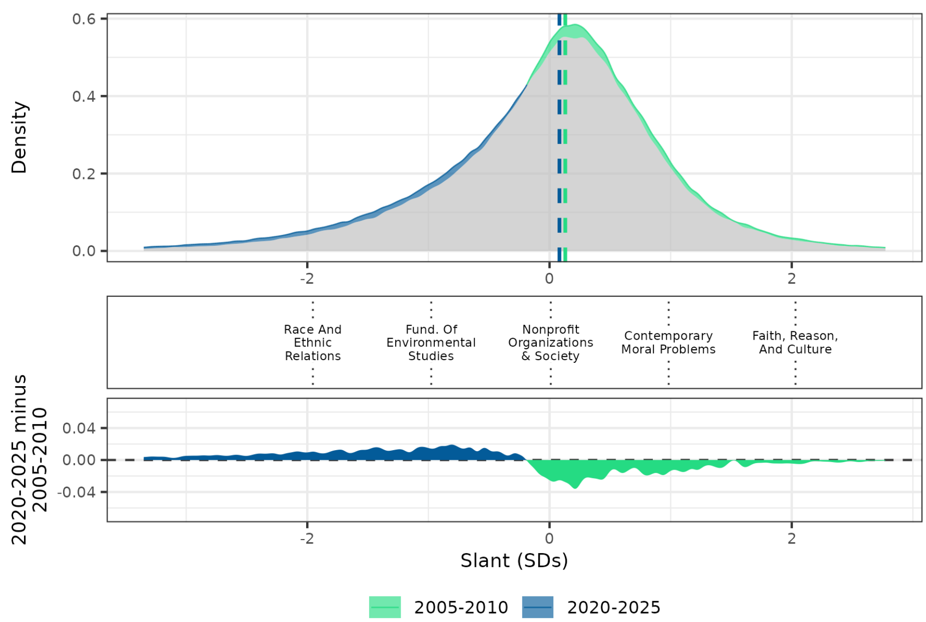
Notes: The second and fourth columns show the 2020-2025 mass below the 2005-2010 percentile cutoff dictated by percentile p in the first column. Two stars represent a change significant at the $p < 0.05$ level; three stars, at the $p < 0.01$ level. The third and fifth columns show the change in mass *relative* to the 2005-2010 baseline. Tail mass changes below the median represent changes in the lower tail; changes above the median represent changes in the upper tail. Analogous to Table B.5, but uses a balanced panel tracking the same 192 schools in both periods.

FIGURE B.3. Distribution of political characteristics for courses 2005-2010 vs. 2020-2025, balanced panel

(A) Political content



(B) Slant



Notes: Each panel compares the distribution of a course content measure — political content (panel a) and slant (panel b) — between the early (2005-2010, green) and late (2020-2025, blue) windows, in 2016-17 cross-sectional standard-deviation units. For slant, positive values are conservative and negative values liberal. Within each panel, the top sub-panel overlays the two kernel density estimates; the dashed vertical lines mark the period-specific medians. The middle sub-panel locates representative courses from Table 1 along the same axis for reference. The bottom sub-panel plots the late-minus-early density difference; positive (blue) regions gained mass and negative (green) regions lost mass. The unit of observation is a course, collapsing contemporaneous sections to a single observation. Figure restricts to a balanced panel of 192 institutions appearing in both periods.

B.2. **Entry and Exit.** Here, we decompose the average trend in course characteristics into courses that enter, exit, and change. The trends in Section 4 could arise from three margins: persistent courses changing their content, new courses entering, and courses exiting. Fix base year $b = 2010$ and target year t.[29] Let Y_b^{all} and Y_t^{all} denote the base-year and year-t catalog means of a standardized outcome (political content or slant, in 2016-SD units). Let Y_b^{surv} and Y_t^{surv} restrict to courses offered in both year t and b, evaluated at their base-year and year-t values. The change in the mean admits the decomposition:

$$\underbrace{Y_t^{\text{all}} - Y_b^{\text{all}}}_{\text{Total change}} = \underbrace{\left(Y_t^{\text{all}} - Y_t^{\text{surv}}\right)}_{\text{Entry}} + \underbrace{\left(Y_t^{\text{surv}} - Y_b^{\text{surv}}\right)}_{\text{Within-course}} + \underbrace{\left(Y_b^{\text{surv}} - Y_b^{\text{all}}\right)}_{\text{Survivor Selection}}. \tag{B.1}$$

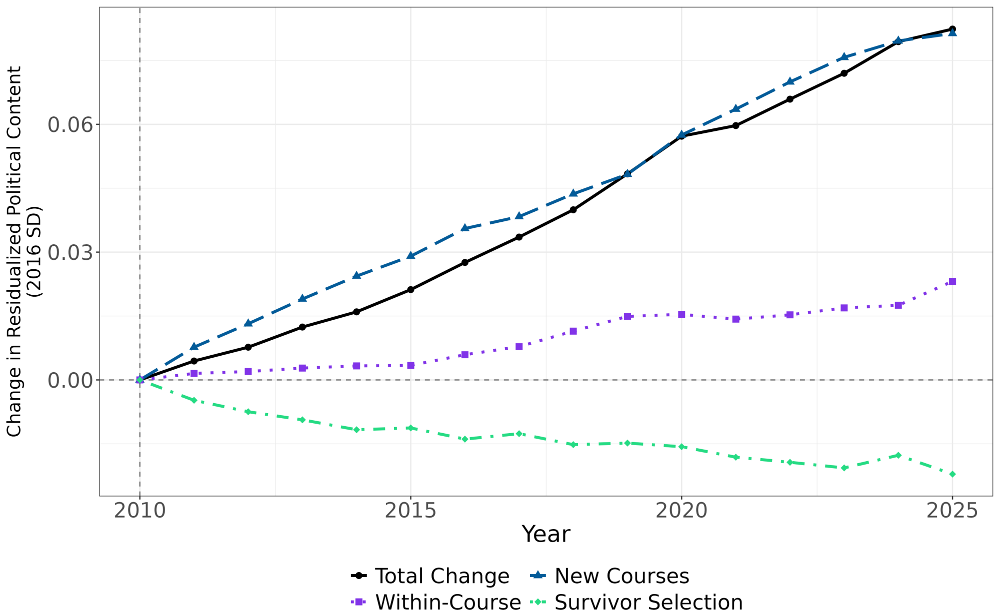
The total change in course characteristics pools both within-course changes and compositional shifts in which courses are offered. The entry term compares all courses in time period t to the set of courses offered in time t that were also offered in time b. The within-course term compares year t to year b for the set of courses that were taught in both years. Finally, the survivor selection term compares the set of courses that survive to year t, evaluated at year b, compared to all courses in year b. We report each component for each target year in 2016-SD units. We compute this both in raw levels and residualized on institution-by-field fixed effects to align with the main text.

Figure B.4 plots the results. For political content, within-course increases in political content are offset by the surviving courses being less political than the average in 2010.[30] Therefore, the total increase in political content closely tracks new courses. This decomposition implies that entering courses are in the right tail of political content — exiting courses may have included more political content than average, but their replacements contained *even more* political content. For slant, we see that surviving courses are typically more conservative than the average course in 2010. There is relatively little change in slant within courses. Finally, new courses are more liberal than the average 2010 course, collectively demonstrating that the aggregate trend in liberal slant comes from the left tail: exiting courses were more liberal than average, but entering courses were even more liberal. In 2025-26, we observe a slight reversion of the entry component, suggesting that the new courses (those offered in 2025-26 but not 2010-11, which comprise courses new in 2025-26 and those introduced in intervening years) have slightly less liberal slant than the new courses in 2024-25.

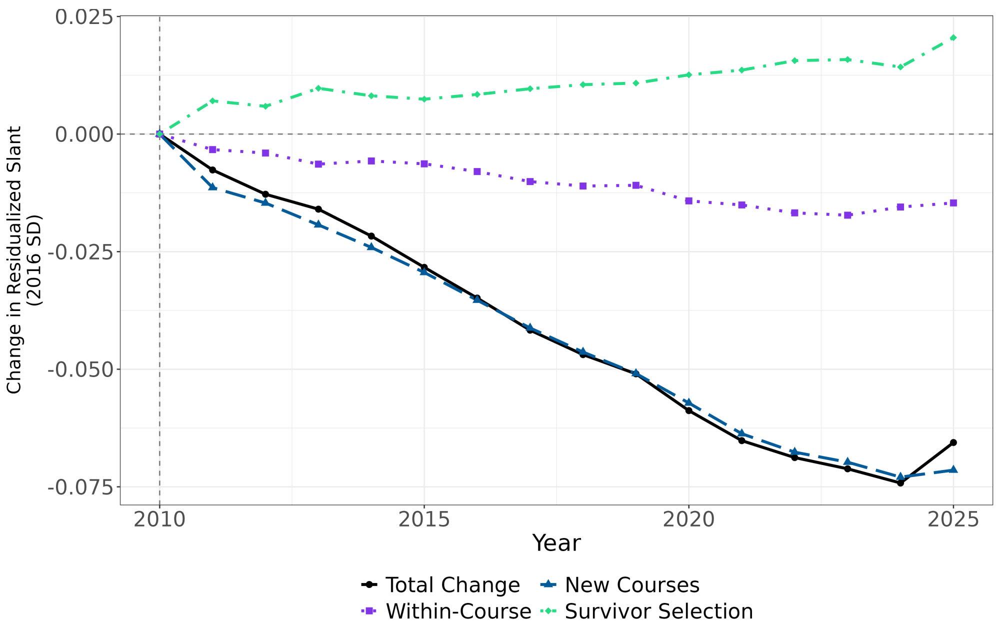
[29]The choice of 2010 as the base year is arbitrary, balancing a longer panel for the decomposition with thicker institutional coverage in later years of the sample. Decomposition results using alternative base years are qualitatively identical.

[30]Figure B.5 plots the survival rate of courses taught in 2010-11.

FIGURE B.4. Entry, exit, and within-course decomposition of trends in political content and slant



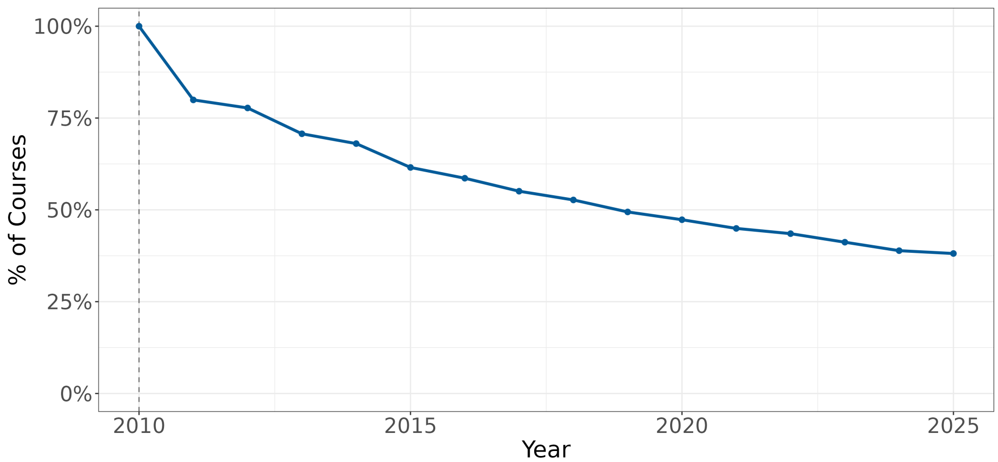
(A) Political content



(B) Slant

Notes: Each panel decomposes the change in the average political content (Panel A) or slant (Panel B) of the course catalog relative to 2010-11 according to Equation B.1. "New courses" capture the contribution of courses not offered in 2010-11 but offered in the indicated year; "survivor selection" captures the contribution of courses offered in 2010-11 but not in the indicated year; "within-course" captures changes arising from courses offered in both 2010-11 and the indicated year. The three components sum to the total change. Outcomes are residualized on institution-by-field fixed effects and expressed in 2016-17 cross-course standard deviation units.

Figure B.5. Survival rate of 2010-11 courses

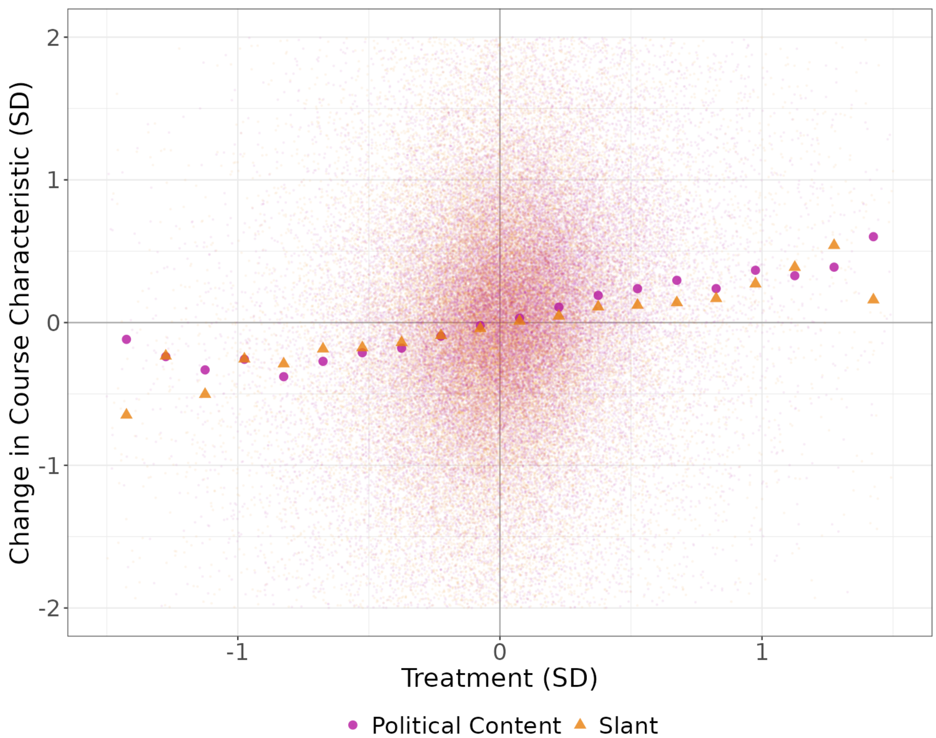


Notes: The figure plots, for each subsequent academic year, the share of courses offered in the 2010-11 base year that continue to appear in the course catalog. A course is counted as surviving when the same course is observed in both 2010-11 and the indicated year. Restricts to institutions continuously observed since 2010-11 or earlier.

Appendix C. Movers Supplement

This appendix contains supplementary material for Section 5, focusing on behavior of instructors who move between institutions.

Figure C.1. On-impact change in course content versus move dosage



Notes: The figure is a binscatter plotting changes in average course political content or slant at the time of a move against the move dosage for that characteristic. Move dosage is the leave-own-courses-out difference between average course characteristics in the instructor's destination and origin institution-field cells at the time of the move. The outcome is the change in the instructor's average course characteristic from the final pre-move period to the first post-move period. The large points report binned means and faint points show the underlying observations. Values are expressed in cross-course standard deviation units.

C.1. **Supplementary Exhibits.**

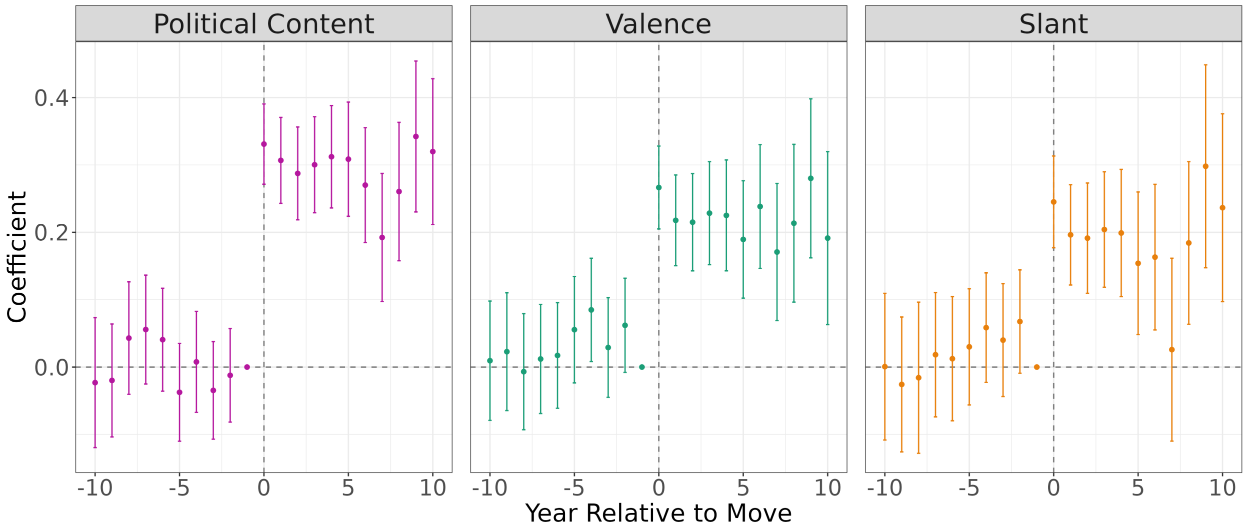
Table C.1. Mover coefficient heterogeneity splits

	Political Content	Slant
Overall	0.375***	0.325***
	(0.013)	(0.015)
Course type		
All courses	0.375***	0.325***
	(0.013)	(0.015)
First-time descriptions only	0.341***	0.187***
	(0.017)	(0.019)
Field group		
STEM	0.323***	0.236***
	(0.028)	(0.029)
Business	0.426***	0.395***
	(0.045)	(0.078)
Arts & Language	0.409***	0.371***
	(0.027)	(0.031)
Humanities	0.396***	0.286***
	(0.040)	(0.038)
Social Sciences	0.404***	0.286***
	(0.036)	(0.039)
Selectivity		
Least Selective	0.459***	0.370***
	(0.027)	(0.031)
Mid Selective	0.298***	0.238***
	(0.028)	(0.035)
Highly Selective	0.191***	0.210***
	(0.047)	(0.075)
Move year		
2008 and earlier	0.264**	0.363***
	(0.111)	(0.134)
2009-2013	0.339***	0.259***
	(0.042)	(0.051)
2014-2018	0.342***	0.308***
	(0.025)	(0.030)
2019-2023	0.384***	0.324***
	(0.022)	(0.024)
2024 and later	0.424***	0.373***
	(0.029)	(0.029)
Institution control		
Public	0.379***	0.325***
	(0.017)	(0.020)
Private	0.250***	0.218***
	(0.042)	(0.051)
State politics		
Red	0.368***	0.287***
	(0.023)	(0.026)
Blue	0.314***	0.370***
	(0.029)	(0.033)

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Notes: The table reports post-move dosage coefficients from the movers specification separately by the indicated instructor or institution characteristics. Because subgroup samples are smaller than in the main analysis, event time is pooled into pre- and post-move periods within five years of the move. The sample is restricted to within-category moves for the characteristic defining each split (for example, highly selective to highly selective institutions). Larger coefficients indicate greater adaptation toward destination course characteristics.

FIGURE C.2. Movers event study, restricted to newly offered course descriptions



Notes: The figure reproduces the event study specification in Figure 7, restricting the sample to course descriptions observed for the first time in the catalog. This restriction limits mechanical adaptation arising because movers are assigned courses with descriptions that already existed at the destination. Points are coefficients on event-time-relative-to-move indicators interacted with move dosage, where dosage is the leave-own-courses-out destination-minus-origin difference within field. Larger coefficients therefore indicate greater adaptation toward destination course characteristics. Standard errors are clustered at the instructor level; vertical bars are 95% confidence intervals.

Table C.2. Within-group mover coefficient equality test

	Political Content				Slant			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Course type								
(1) All courses								
(2) First-time only	-0.035*				-0.138***			
	(0.019)				(0.022)			
Field group								
(1) STEM								
(2) Business	0.102*				0.159*			
	(0.052)				(0.083)			
(3) Arts & Language	0.085**	-0.017			0.135***	-0.024		
	(0.039)	(0.052)			(0.043)	(0.084)		
(4) Humanities	0.073	-0.029	-0.012		0.050	-0.109	-0.085*	
	(0.048)	(0.060)	(0.048)		(0.048)	(0.087)	(0.049)	
(5) Social Sciences	0.081*	-0.021	-0.004	0.008	0.050	-0.109	-0.085*	0.000
	(0.045)	(0.057)	(0.045)	(0.054)	(0.049)	(0.087)	(0.050)	(0.055)
Selectivity								
(1) Least Selective								
(2) Mid Selective	-0.161***				-0.132***			
	(0.039)				(0.047)			
(3) Highly Selective	-0.268***	-0.107**			-0.160**	-0.027		
	(0.054)	(0.055)			(0.081)	(0.083)		
Move year								
(1) 2008 and earlier								
(2) 2009-2013	0.075				-0.104			
	(0.118)				(0.142)			
(3) 2014-2018	0.079	0.004			-0.055	0.049		
	(0.113)	(0.049)			(0.136)	(0.059)		
(4) 2019-2023	0.120	0.045	0.042		-0.039	0.065	0.016	
	(0.112)	(0.047)	(0.033)		(0.135)	(0.057)	(0.039)	
(5) 2024 and later	0.160	0.085*	0.082**	0.040	0.010	0.114*	0.065	0.049
	(0.114)	(0.051)	(0.038)	(0.036)	(0.136)	(0.059)	(0.042)	(0.038)
Institution control								
(1) Public								
(2) Private	-0.130***				-0.106*			
	(0.045)				(0.055)			
State politics								
(1) Red								
(2) Blue	-0.054				0.083**			
	(0.037)				(0.042)			

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Notes: Tests of whether the heterogeneous coefficients from Table C.1 are different within-group. Each cut type has a lower-triangular matrix comparing the coefficient in row i to that of column j. The left set of columns tests for political content, the right for slant.

Table C.3. Mover coefficient heterogeneity splits, new descriptions only

	Political Content	Slant
Overall	0.341***	0.187***
	(0.017)	(0.019)
Field group		
STEM	0.314***	0.164***
	(0.040)	(0.040)
Business	0.330***	0.118
	(0.068)	(0.089)
Arts & Language	0.294***	0.237***
	(0.045)	(0.047)
Humanities	0.322***	0.186***
	(0.043)	(0.048)
Social Sciences	0.397***	0.054
	(0.042)	(0.042)
Selectivity		
Least Selective	0.279***	0.206***
	(0.036)	(0.040)
Mid Selective	0.350***	0.212***
	(0.045)	(0.044)
Highly Selective	0.183***	0.181***
	(0.045)	(0.063)
Move year		
2008 and earlier	0.424***	0.175
	(0.123)	(0.122)
2009-2013	0.350***	0.174***
	(0.050)	(0.066)
2014-2018	0.307***	0.101***
	(0.031)	(0.036)
2019-2023	0.345***	0.221***
	(0.029)	(0.032)
2024 and later	0.370***	0.234***
	(0.037)	(0.039)
Institution control		
Public	0.332***	0.187***
	(0.024)	(0.026)
Private	0.212***	0.170***
	(0.038)	(0.047)
State politics		
Red	0.319***	0.151***
	(0.034)	(0.036)
Blue	0.284***	0.174***
	(0.036)	(0.039)

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Notes: The table reproduces the heterogeneous movers estimates in Table C.1, restricting the sample to course descriptions observed for the first time in the catalog. Because subgroup samples are smaller than in the main analysis, event time is pooled into pre- and post-move periods within five years of the move. The sample is restricted to within-category moves for the characteristic defining each split (for example, highly selective to highly selective institutions). Larger coefficients indicate greater adaptation toward destination course characteristics.

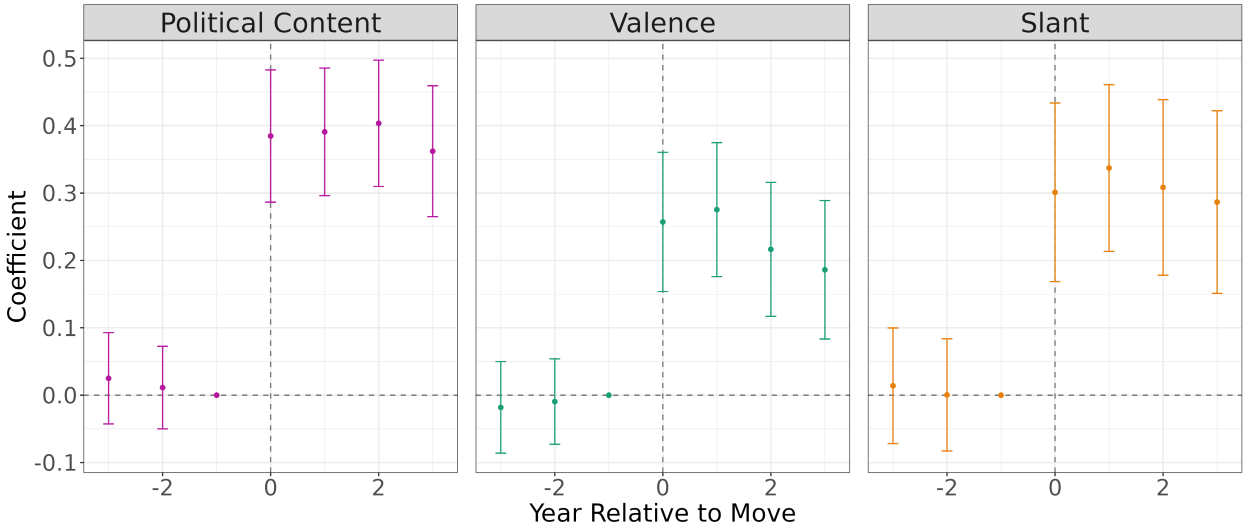
Table C.4. Within-group mover coefficient equality test, new courses only

	Political Content				Slant			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Field group								
(1) STEM								
(2) Business	0.016				-0.046			
	(0.079)				(0.097)			
(3) Arts & Language	-0.020	-0.035			0.073	0.119		
	(0.061)	(0.081)			(0.062)	(0.100)		
(4) Humanities	0.008	-0.008	0.027		0.022	0.068	-0.051	
	(0.059)	(0.080)	(0.062)		(0.062)	(0.100)	(0.067)	
(5) Social Sciences	0.083	0.068	0.103*	0.076	-0.111*	-0.064	-0.183***	-0.133**
	(0.059)	(0.080)	(0.062)	(0.060)	(0.058)	(0.098)	(0.063)	(0.064)
Selectivity								
(1) Least Selective								
(2) Mid Selective	0.070				0.005			
	(0.058)				(0.060)			
(3) Highly Selective	-0.096*	-0.167***			-0.026	-0.031		
	(0.058)	(0.064)			(0.074)	(0.077)		
Move year								
(1) 2008 and earlier								
(2) 2009-2013	-0.074				-0.001			
	(0.129)				(0.136)			
(3) 2014-2018	-0.117	-0.043			-0.074	-0.073		
	(0.123)	(0.058)			(0.124)	(0.075)		
(4) 2019-2023	-0.079	-0.005	0.038		0.046	0.047	0.120**	
	(0.123)	(0.057)	(0.042)		(0.123)	(0.073)	(0.048)	
(5) 2024 and later	-0.054	0.020	0.063	0.026	0.059	0.060	0.133**	0.013
	(0.125)	(0.062)	(0.048)	(0.047)	(0.125)	(0.077)	(0.053)	(0.050)
Institution control								
(1) Public								
(2) Private	-0.120***				-0.017			
	(0.045)				(0.053)			
State politics								
(1) Red								
(2) Blue	-0.035				0.023			
	(0.049)				(0.054)			

$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Notes: Tests of whether the heterogeneous coefficients from Table C.3 are different within-group. Each cut type has a lower-triangular matrix comparing the coefficient in row i to that of column j. The left set of columns tests for political content, the right for slant. Analogous to Table C.2, but using only courses taught for the first time.

Figure C.3. Movers event study, balanced-panel instructors



Notes: The figure reproduces the event study specification in Figure 7, restricting the sample to instructors observed teaching in each of the three years immediately before and after the move. Points are coefficients on event time indicators interacted with move dosage, defined as the leave-own-courses-out destination-minus-origin difference in the corresponding course characteristic within field. Larger coefficients indicate greater adaptation toward destination course characteristics. Standard errors are clustered at the instructor level; vertical bars are 95% confidence intervals.

C.2. **Matching Procedure.** Our movers design requires identifying faculty across different university catalogs which may have different formatting standards. To do so, we use an iterative matching procedure where we start with the strictest possible match, then gradually relax constraints to identify further matches.

The first step is to standardize formatting to permit cross-university comparisons. This requires digesting, for example, "J. D. Light," "Dr. Jacob Light," and "Light, Jacob" into a common format. We do this via a combination of regular expressions and manual correction, eventually extracting–for each institution–a "first name" and "last name" field, and for some institutions a "middle initial" field. We apply a similar standardization procedure to field names, aligning "English Literature" with "English." We omit any instructors teaching in a field labeled as "Other."[31]

By starting with the highest-quality matches and gradually relaxing our constraints, we ensure lower-quality matches never crowd out higher-quality ones. In all cases, we bar matches across universities within the same year, such that the same faculty member never has two appointments at the same time to avoid matching common names. We start by matching on first name, middle initial, last name, and field–e.g., "Gideon," "S," "Moore," in "Economics." On the second pass, we then match on first name, last name, and field, requiring there only be no *contradictory* middle initials; in this case, "Gideon S. Moore in Economics" would match with "Gideon Moore in Economics," but not "Gideon J. Moore in Economics." We omit placeholder instructor names–"TBA," "Pending," and the like. Note in all cases we match on field; this means we omit instructor moves across fields (e.g., economics to public policy). We believe this to be a worthwhile trade-off; the number of false positives introduced by relaxing this constraint swamps the number of identified true cross-field moves.

C.3. **Bargaining Model.** Suppose a course has N stakeholders–instructors, students, institutions, donors, alumni, and the like. Each stakeholder n has ideal point x_n, disagreement surplus D_n, and weight ω_n such that $\sum \omega_n = 1$. Formally, D_n is the surplus from reaching an agreement at ideal point x_n, relative to disagreement; if $d_n \leq 0$ is the disagreement payoff to n, and utility of an agreement at x_n is 0, then $D_n = -d_n$. With quadratic loss away from their ideal point, we can

[31] Typically non-standard university institutes or other non-academic disciplines. This restriction eliminates observations in institution-specific institutes or other non-standard academic fields.

write the joint Nash problem among all stakeholders as:

$$s^*_{ijt} = arg\max_s \prod_{n=1}^{N} (D_n - (s - x_n)^2)^{\omega_n}$$

Logging the objective and maximizing yields the following first-order condition:

$$\sum_n \frac{-2\omega_n(s^* - x_n)}{D_n - (s^* - x_n)^2} = 0$$

If D_n are sufficiently large, the distance term in the denominator becomes negligible. This requires, in some sense, that disagreement is "sufficiently bad" that a bargain will always occur. We think this is reasonable in our setting–faculty generally must teach *some* course to retain employment, students *must* enroll in *some* course to receive credit, etc. Under this large-D_n approximation, we get:

$$\sum_n \frac{\omega_n}{D_n}(s - x_n) = 0$$

Note in this case the bargaining weight ω_n is modulated by disagreement point D_n to create an *effective* bargaining weight $\tilde{\omega}_n := \frac{\omega_n/D_n}{\sum_k \omega_k/D_k}$, allowing us to re-write this expression as:

$$\sum_n \tilde{\omega}_n(s - x_n) = 0$$

Separating out the instructor stakeholder term and substituting ϕ for the instructor ideal point and (1 - α) for the bargaining weight, we get:

$$(1 - \alpha)(s - \phi) + \sum_{n \neq instr} \tilde{\omega}_n(s - x_n) = 0$$

Defining a "context ideal point" $\kappa = \frac{1}{\alpha}\sum_{n \neq instr} \tilde{\omega}_n x_n$ generated from weighting the individual ideal points by the bargaining weights, we can re-write the second term of this expression as:

$$\begin{aligned}
\sum_{n \neq instr} \tilde{\omega}_n(s - x_n) &= \sum_{n \neq instr} \tilde{\omega}_n s - \sum_{n \neq instr} \tilde{\omega}_n x_n \\
&= s \sum_{n \neq instr} \tilde{\omega}_n - \sum_{n \neq instr} \tilde{\omega}_n x_n \\
&= s(1 - (1 - \alpha)) - \alpha\kappa
\end{aligned}$$

$$= \alpha(s - \kappa)$$

Thus, we can define this term as a difference between equilibrium slant s and the context ideal point κ, multiplied by the non-instructor bargaining weight α.

Plugging this back into our first-order condition and rearranging recovers the linear expression for equilibrium slant we use in our estimating equation:

$$s^* = (1 - \alpha)\phi + \alpha\kappa$$

Holding instructor ideal points ϕ fixed around a move, letting κ_{t-1} denote the environment pre-move and κ_t be the environment post move, we have:

$$\begin{aligned} s_t^* - s_{t-1}^* &= [(1 - \alpha)\phi + \alpha\kappa_t] - [(1 - \alpha)\phi + \alpha\kappa_{t-1}] \\ &= \alpha(\kappa_t - \kappa_{t-1}) \end{aligned}$$

Exactly the reduced form estimating equation, using $\delta_j = \kappa_t - \kappa_{t-1}$. Note that in this framework, the ideal point of the institution includes non-moving instructor ideal points.

Appendix D. Demand Supplement

This appendix contains supplementary material for Section 6, focusing on trends in enrollment and student demand.

D.1. Computation of Market Shares.

We compute market shares from realized enrollment rather than potential enrollment. Given the market fixed effect, the two are numerically equivalent. Assume that X_{jt} includes the within nest share, so this model is nested logit. From the Berry (1994) inversion,

$$\ln(\tilde{p}_{jt}) - \ln(p_{0\,i(j)t}) = \beta X_{jt} + \xi_{jt}$$

Expanding

$$\ln(\tilde{p}_{jt}) = \ln(\text{enrollment}_{jt}) - \ln(\text{total possible}_{i(j)t})$$

and adding

$$\ln(\text{total possible}_{i(j)t}) + \ln(p_{0\,i(j)t})$$

while subtracting

$$\ln(\text{total enrolled}_{i(j)t})$$

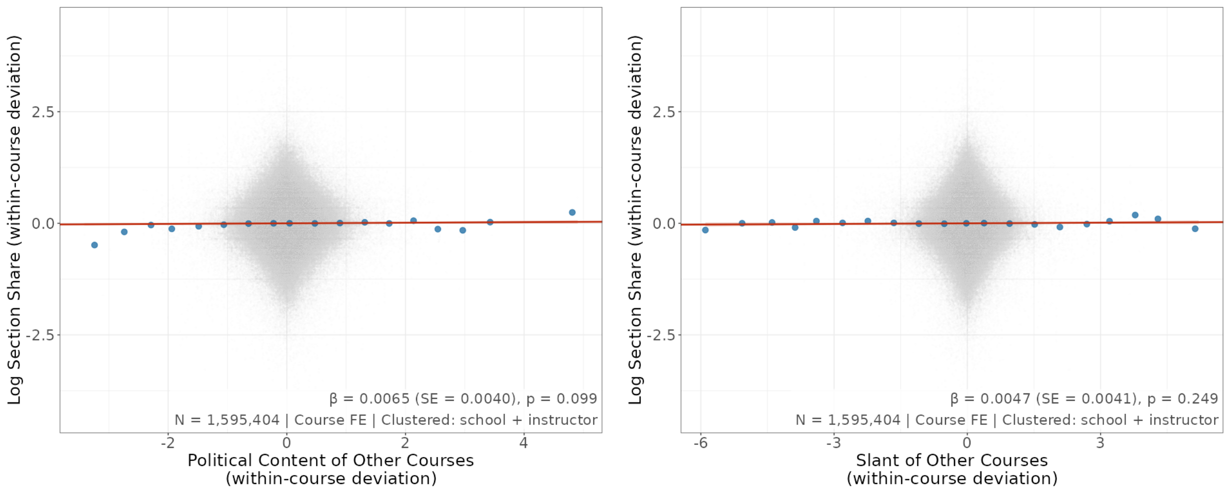
on both sides yields

$$\ln(p_{jt}) = \beta X_{jt} + \underbrace{\ln(p_{0\,i(j)t}) + \ln(\text{total possible}_{i(j)t}) - \ln(\text{total enrolled}_{i(j)t})}_{\kappa_{i(j)t}} + \xi_{jt}$$

so the market fixed effect reproduces the standard nested logit format.

D.2. Supplementary Exhibits.

Figure D.1. Instructor desirability versus the political characteristics of their other courses

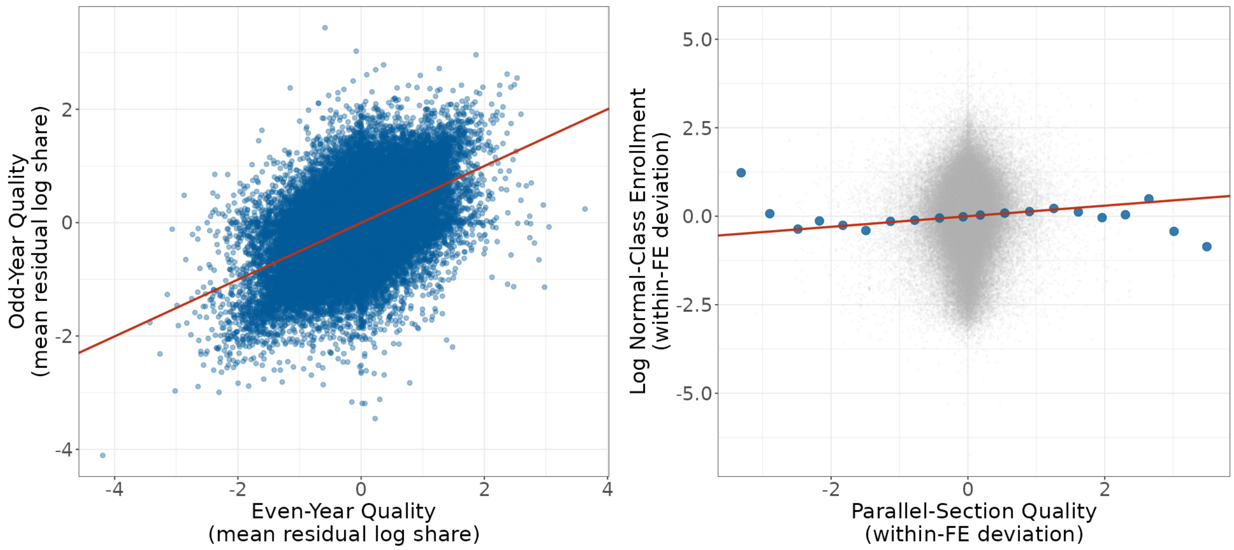


(A) Political content

(B) Slant

Notes: Each panel tests whether the political characteristics of an instructor's courses are related to a measure of that instructor's desirability. Instructor desirability is estimated using residual enrollment differences between sections of the same course offered at the same institution and time but taught by different instructors, thereby holding the shared course description fixed. The horizontal axis reports the average political content (Panel A) or slant (Panel B) of the instructor's other courses. The vertical axis reports instructor desirability as proxied by excess demand in these shared courses. Faint points show instructor-level observations, binned means summarize the relationship, and the line reports the linear fit. The sample applies the four-year instructor history restriction used in the IV demand analysis. Instructors who teach shared courses account for 56.0% of instructors in the IV sample.

Figure D.2. Validation of the residual enrollment-based instructor desirability measure



(a) Even/Odd Split (b) Enrollment

Notes: The figure validates the residual-enrollment measure of instructor desirability constructed from courses with multiple parallel sections. Panel A compares instructor desirability estimated using even academic years with the corresponding estimate using odd academic years; the relationship tests whether the measure captures a persistent instructor component rather than transitory enrollment noise. Panel B relates desirability estimated from shared, multi-section courses to residual log enrollment in the same instructors' courses that do not have parallel sections, providing an out-of-sample test of whether the measure predicts student demand. The regressions used to residualize enrollment in Panel (b) control for institution-by-field and academic term.

Table D.1. First-stage estimates for the student demand instruments

	Log Inside Share	Political Content	Slant
Log Num. Dept. Courses	-0.9460***	0.0036	0.0188
	(0.0091)	(0.0781)	(0.0749)
Political Content (4-yr instr. lag)	0.0269***	0.6213***	-0.0015
	(0.0070)	(0.1350)	(0.0849)
Slant (4-yr instr. lag)	-0.0213***	0.0019	0.6248***
	(0.0061)	(0.0710)	(0.1750)
Year-interacted instruments	Yes	Yes	Yes
Conditional F (KP/SW, cluster)	28448.4	15502.6	16635.2
KP rk Wald F (joint, full system)		250.8	
Observations		8,394,204	

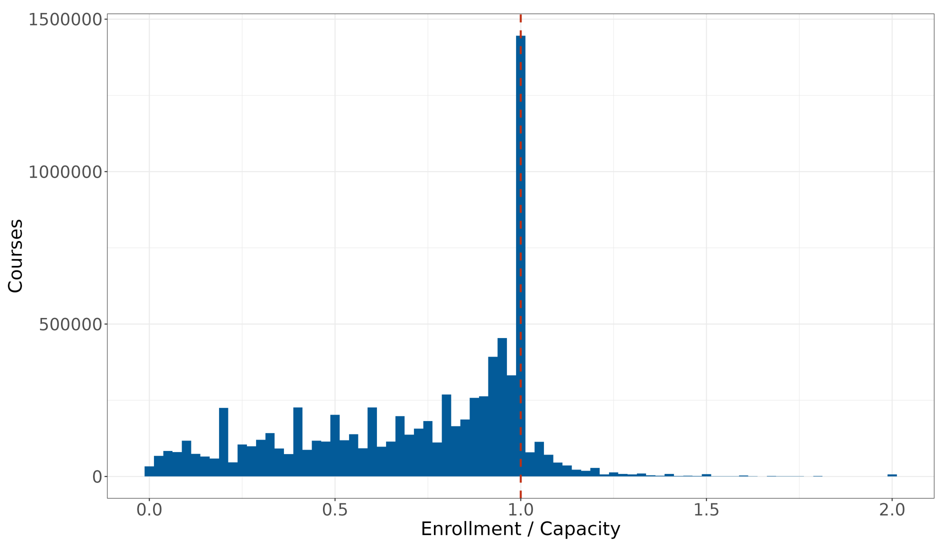
Market (school × year × sem), course-level, school × field, and field × year FEs; SEs clustered by school.
Main-instrument coefficients reflect the reference-year (2016) first-stage relation;
year-specific deviations load on the unreported year-interacted instruments.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Notes: The table reports first-stage estimates and diagnostics for the instrumented nested logit demand system. The excluded instruments are the average political content and slant of the same instructor's courses four years earlier and the log number of course offerings in the field. For compactness, only the 2016 first-stage coefficients are shown.

D.3. **Course Capacity.** Here, we provide summary statistics and robustness checks on course capacities. First, in Figure D.3, we show the distribution of course enrollment divided by course capacity. In some cases, the course enrollment exceeds 100%, meaning that course caps are not binding. On average, 19% of courses are at capacity exactly, while 6% of courses are over their nominal caps. The fact that a non-trivial fraction of courses with binding capacities exceed the nominal cap suggests that course caps do not play too big a role in student choice sets.

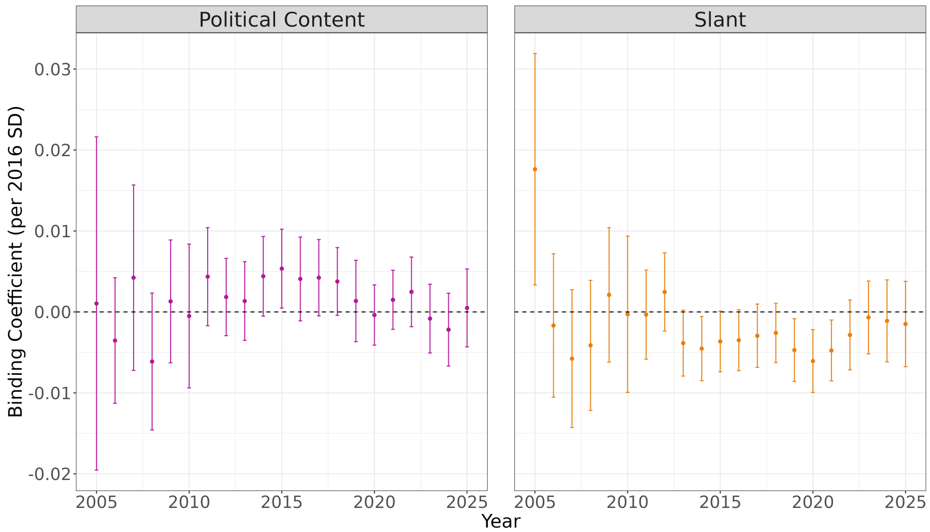
Second, in Figure D.4, we re-estimate the demand system, but instead predict whether the course is at capacity. We find that political characteristics (in the context of the demand system) do not meaningfully predict the probability that the course cap binds.

Figure D.3. Distribution of course fill rates



Notes: The figure plots the distribution of course fill rates, defined as observed enrollment divided by reported course capacity, among courses for which capacity is observed. A value of one indicates enrollment equal to the nominal capacity. Values above one occur when enrollment exceeds the reported capacity, indicating that the recorded capacity does not always represent a hard enrollment constraint.

Figure D.4. Political characteristics and the probability that a course fills to capacity



Notes: Estimates come from separate regressions of an indicator for whether course enrollment equals course capacity on time indicators interacted with course political content (left) and slant (right), respectively. Regressions include the same fixed effects as in the demand model and use the same course sample restrictions as the IV demand analysis. Observations at the course level. Standard errors are clustered at the institution level; vertical bars indicate 95% confidence intervals.

Appendix E. External Ratings

This appendix contains supplementary material detailing human and LLM-based validation of our measures.

We solicit Likert scale ratings of political content and slant from both human volunteers and LLMs. Both groups' assessments of courses align with our embedding measure on both political content and slant.

Given this benchmark, why not use these Likert ratings for the whole analysis in place of the embedding approach? We prefer our embedding-based approach to LLM annotations for the main analysis for three reasons. First, Likert ratings are potentially sensitive to the scoring guide, introducing researcher discretion at a different point in the pipeline. Second, embedding-based measures are continuous, whereas Likert ratings compress the distribution into five integer values. Third, generating annotations at the scale of our full dataset would be cost-prohibitive; even the LLM ratings for this validation subsample alone cost substantially more than embedding the complete corpus.

E.1. **Sample Construction.** We select 5,000 courses in our data to validate our measure against external labeling. Given the vast majority of courses in our sample are neither political nor slanted, this sample is stratified by our embedding measure to ensure our validation covers the whole range of potential scores. In particular, we first construct three political content strata:

- 0th to 50th percentile
- 50th to 95th percentile
- 95th to 100th percentile

Within the top political content stratum, we further stratify into three slant strata:

- 0 to 5th percentile
- 5th to 95th percentile
- 95th to 100th percentile

In total this gives us five strata: non-political classes, moderately political classes, highly political liberal classes, highly political neutral classes, and highly political conservative classes. We sample equally from these five strata to ensure full coverage of the range of our embedding measure.

Due to changes in our cleaning algorithm subsequent to the validation exercise, we need to map participant responses to updated descriptions. The large majority (71.7%) of the descriptions experience no non-punctuation changes. Another 19.2% are uniquely mapped between cleaning regimes by extreme containment or Jaccard similarity. Of the remaining 9.1%, we have an LLM examine the most similar courses by Jaccard similarity and it selects a high-confidence match; this recovers an additional 7.8% of the sample. The remaining 1.3% of courses could not be confidently matched to a current description, so are omitted from the validation exercise as the text the raters reviewed is no longer comparable to the text post-cleaning (e.g., the whole description had been boilerplate that is now scrubbed).

E.2. **Scoring Guide.** To normalize our raters and be concrete about our meanings of "political content" and "slant," we develop a Likert scoring guide. This defines criteria for courses to be scored on a one-to-five scale for both political content and slant. The scoring guide is reproduced in its entirety at the end of this section.[32] We provide anchors to define each of the 1-to-5 bins

[32]In a previous version of this project, we referred to the construct we currently call the amount of political content as a text's "politicization." This term was operative at the time we conducted the validation. For transparency, we reproduce the scoring guides exactly as respondents saw them, rather than updating with the new terminology.

for both measures. For each bin, we also provide worked examples including both sample course descriptions and rationales for their scores.

To preserve instructor privacy, we do not reproduce "true" course descriptions used for scoring in the public version of the scoring guide. Rather, we submit the true description and scoring rationale to ChatGPT version 5.5 with the below prompt. In the scoring guide below, we have replaced the true descriptions with their "synthetic" counterparts.

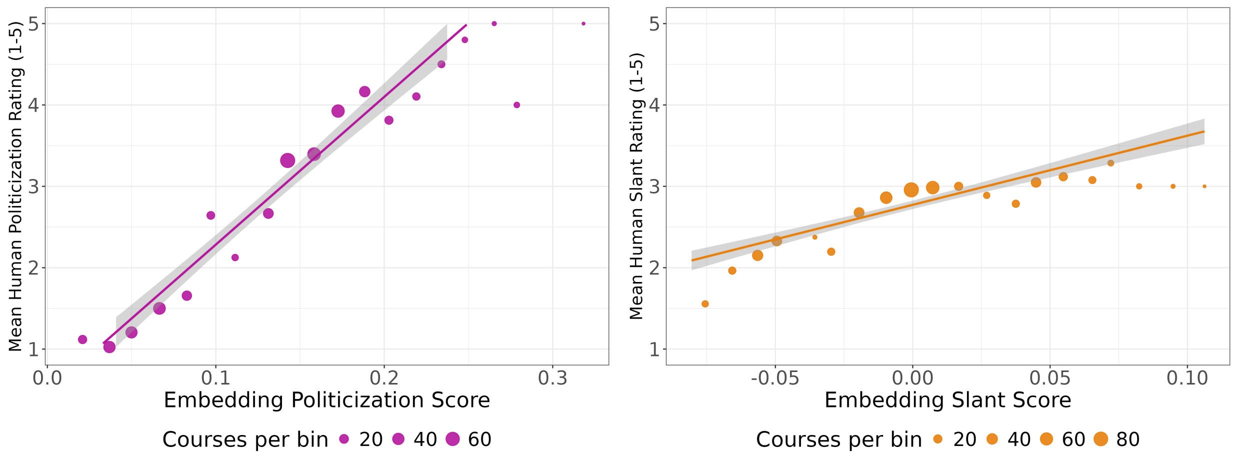
```
You paraphrase university course descriptions so they can be shown in a public
methodology document without revealing which specific course was the source.

Your output must:
1. Preserve every rhetorical, topical, and stylistic feature that the supplied
rationale relies on. If the rationale calls out specific words, framings,
advocacy language, or ideological markers, the paraphrase must still exhibit
them (using different wording).
2. Remove anything that could let a reader reverse-search the original course:
proper nouns, named authors, named programs or centers, distinctive course
titles, course numbers, unique phrases, geographic specifics, and references
to particular institutions or general-education area codes.
3. Match the original in length and register (academic catalog prose,
lowercase if the original is lowercase, no markdown).
Respond with JSON of the form {paraphrase: ...} and nothing else.
```

E.3. **Human Scoring.** Human hand-labels are strongly correlated with our embedding measures. We solicited four volunteer research assistants from the Stanford community[33] to label courses according to the coding guide.

Volunteers received a half hour of training at the start of the session. To ensure adherence to the scoring guide, we walked through each potential score individually and worked through the given example. Participants were then given a set of five "practice questions" the authors agreed on scores for ex-ante; raters worked through the practice exercises individually and then worked with the authors in a group discussion to understand how the scoring guide would categorize these examples.

[33]Authors solicited volunteers from former students and pre-doctoral research assistants at Stanford. No assistant has or has ever had an employment or personal relationship with any of the authors. Nothing in the solicitation suggested a connection to political slant.

Figure E.1. Human Likert ratings versus embedding-based measures



(A) Political content (B) Slant

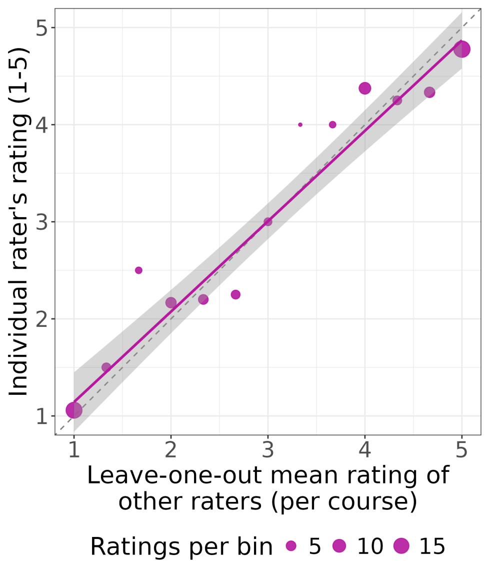
Notes: Each panel plots embedding-based scores (horizontal axis) against human Likert ratings (vertical axis) for political content (left) and slant (right). Spearman correlations are 0.71 for political content and 0.48 for slant.

On completion of the training, volunteers and authors both scored examples for two hours, taking twenty-minute shifts with five-minute breaks. The first twenty courses for each respondent were the same, allowing us to check for agreement among the respondents. Subsequent courses were all drawn at random from the stratified sample and graded by one rater each. In total this gives us twenty courses scored by all reviewers and 1,200 courses scored by one reviewer each.[34] Volunteers were invited to a presentation of preliminary results from the project after completion, but were not obligated to stay.

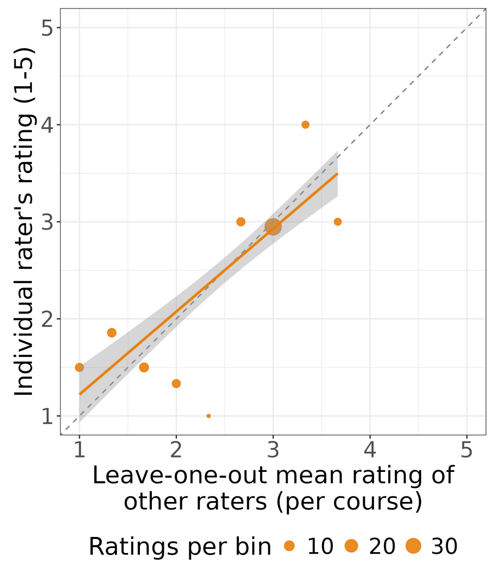
Human scores are consistent with our embedding scores, as seen in Figure E.1. We see strong correlation between measures for both political content (Spearman correlation = 0.71) and slant (Spearman correlation = 0.48). We attribute the shallower slope of the slant estimates to more severe measurement error, both philosophically and mechanically. It is more difficult to judge the "way" something is talked about than "what" is talked about, making this a task more prone to error at baseline; moreover, the differential similarity nature of our slant measure makes it susceptible to measurement error in both Democratic and Republican similarity, while political content is subject only to error in identifying the politically engaged topics. Raters exhibit significant consensus when

[34] This does not include courses scored by the authors, nor do the results below. Authors scored courses as a quality check. Results including and excluding author scores are substantively unchanged.

FIGURE E.2. Inter-rater agreement in human Likert ratings



(A) Political content



(B) Slant

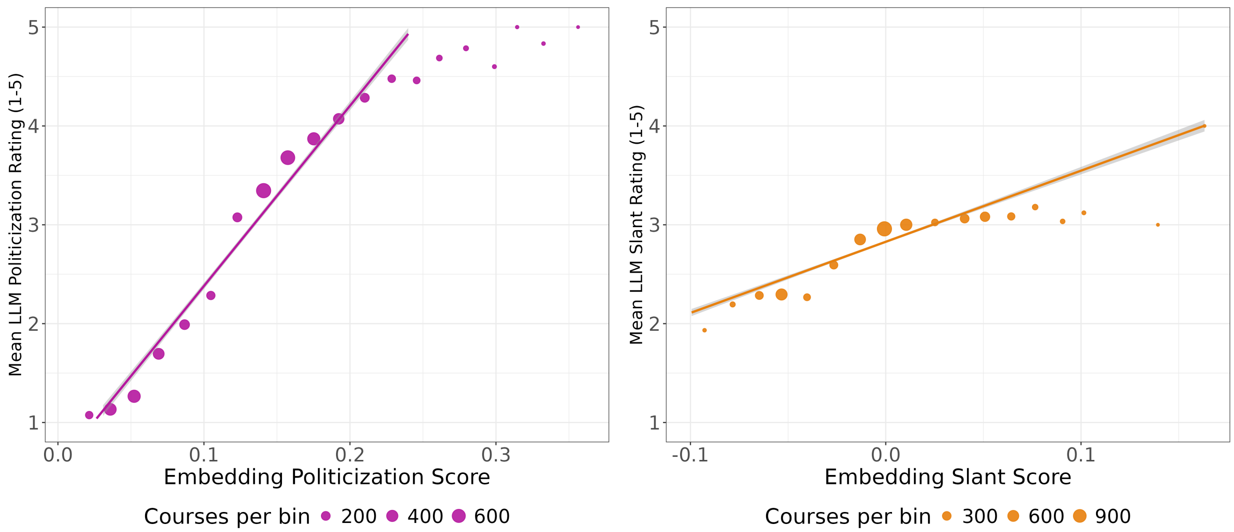
Notes: Each panel plots a course's leave-own-out average human rating (horizontal axis) against an individual rater's rating (vertical axis) for political content (left) and slant (right). Perfect agreement lies on the 45-degree line; pure noise would produce a horizontal line. Intra-class correlation coefficients are 0.83 for political content and 0.68 for slant.

scoring their overlapping courses as seen in Figure E.2, achieving an intra-class correlation coefficient of 0.83 for political content and 0.68 for slant.

E.4. **LLM Scoring.** LLM-based scores are similarly strongly correlated with our embedding measures. For each course in the stratified validation sample we ask `gpt-5.4-mini` to rate each description from 1 to 5 on political content and slant according to the same scoring guide used for human validation. We solicit these scores three times to check for uncertainty.

Our embedding-based measures align closely with the LLM annotations, as seen in Figure E.3. Spearman correlations are 0.73 for political content and 0.53 for slant, extremely similar to the human labels. The LLM produces highly consistent ratings across replications — 88% of political content ratings and 94% of slant ratings are identical across iterations.

Figure E.3. LLM Likert ratings versus embedding-based measures



(a) Political content

(b) Slant

Notes: Each panel plots embedding-based scores (horizontal axis) against LLM (`gpt-5.4-mini`) Likert ratings (vertical axis) for political content (left) and slant (right). Spearman correlations are 0.73 for political content and 0.53 for slant.

Scoring Guide for Course Ratings

Anchor descriptions and worked examples for the 1-5 Likert scales used to rate course descriptions on politicization and slant. Examples were manually selected, rated, and annotated by the researcher.

Section 1: Politicization Scale (1–5)

Anchor Definitions

Rating	Description
1	Not political: Course content is technical, scientific, artistic, or otherwise has no connection to political institutions, policy, ideology, or social movements.
2	Slightly political: Course is primarily focused on non-political subject matter but touches on topics with political dimensions (e.g., a science course that mentions environmental policy, or a communications course covering intercultural dynamics).
3	Moderately political: Course engages substantively with politically relevant topics — such as social institutions, inequality, or cultural conflict — but politics is not the organizing frame of the course.
4	Substantially political: Course is centrally organized around political topics, institutions, or conflicts, even if it also covers non-political material.
5	Highly political: Course is explicitly and primarily about political processes, ideologies, policy, governance, or law.

Worked Examples

Example 1 **(Rating: 1)**

Course Description:

infinite-dimensional algebraic structures and mathematical physics i: introduction to the infinite-dimensional algebraic systems that appear in contemporary mathematics and mathematical physics: oscillator and conformal symmetry algebras, representations of oscillator algebras, highest-weight modules for conformal symmetry algebras, determinant formulas for reducibility, and unitary and discrete-series representations.

Rationale: This course is entirely about mathematics with no political content.

Example 2 **(Rating: 2)**

Course Description:

human communication: a survey of human communication, covering language acquisition, spoken, nonspoken, one-to-one, small-team, workplace, public, and cross-cultural communication. this course may be applied toward a general education communication requirement.

Rationale: Communication is a social topic that could touch on political dimensions (e.g., public and intercultural communication), but the course is focused on communication skills and theory.

Example 3 (Rating: 3)

Course Description:

overview of aging services: reviews concerns involved in supporting older adults, their relatives, and those who provide care. considers biological, psychological, and sociocultural dimensions of later life, along with distinctive challenges in delivering services to older people. analyzes pressures affecting older individuals, and addresses key issues connected to longer life expectancy, changing families, institutional care, and the place of older adults in society.

Rationale: While focused on a specific topic — providing support for older adults — the course touches on several social issues that are politically relevant such as institutionalization and the role of older persons in society.

Example 4 (Rating: 4)

Course Description:

race/gender and intergroup relations: this course examines the sociological and social-psychological aspects of racial relations, ethnic contact, and the evolving position and standing of women. It emphasizes a domestic setting while also considering issues affecting women and minority populations in other regions and their significance for global affairs. It further reviews what scholars in sociology and social psychology have found about ways to improve relations between majority and marginalized groups.

Rationale: This course is political in that the government often regulates legal protections by demographic, and the course discusses policies for improving relations. However, the course itself is not about politics per se.

Example 5 (Rating: 5)

Course Description:

energy and climate coordination across a regional bloc: current political dynamics in leading fossil fuel producers across multiple subregions. the potential for developing sustainable alternative energy resources for export, particularly biofuels, and its impact on agricultural policy, environmental protection, and food prices. the feasibility of creating regional energy security networks such as proposed cross-border security and prosperity arrangements.

Rationale: While not explicitly about any country's political system, this course is clearly 'political', insofar as it focuses on the role of the government in regulating energy production to address climate change.

Section 2: Slant Scale (1–5)

Anchor Definitions

Rating	Description
1	Strong progressive slant: The course has a clear and intense progressive slant. The description actively participates in or advocates for a left-aligned project — e.g., using language of resistance, liberation, or social transformation, or treating progressive normative claims as premises rather than subjects of inquiry.
2	Moderate progressive slant: The course has a detectable progressive slant. This may come from its choice of subject matter (e.g., inequality, identity, environmental justice, critical perspectives on power), its framing (e.g., a neutral topic described with progressive vocabulary), or both — but the course is not overtly activist in posture.
3	Neutral, balanced, or apolitical: No directional slant is detectable. The course either covers non-ideological material, presents balanced perspectives, or engages with political topics without signaling a left-right orientation. A course on a political subject (e.g., American government, international relations) still receives a 3 if the description does not imply a slant.
4	Moderate conservative slant: The course has a detectable conservative slant. This may come from its choice of subject matter (e.g., the moral case for free markets, constitutional originalism, natural rights and ordered liberty, the Western canon as civilizational foundation, traditional family and sexual ethics), its framing (e.g., a neutral topic described with conservative vocabulary — 'heritage,' 'foundations,' 'undermined'), or both — but the course maintains an academic rather than devotional or advocacy posture.
5	Strong conservative slant: The course has a clear and intense conservative slant. The description actively participates in or advocates for a right-aligned project — e.g., treating biblical authority or natural rights as foundations for public policy, framing traditional institutions as under threat, presenting conservative normative claims as premises rather than subjects of inquiry, or using language of decline, restoration, or defense of heritage.

Worked Examples

Example 1 **(Rating: 1)**

Course Description:

This course examines the afterlife of confessional poetry by women and queer, trans, and nonbinary writers. Often dismissed as narcissistic, the poets considered here deploy insurgent self-exposure to unsettle the social and legal construction of gender and sexuality as "personal" matters outside the realm of political concern. In doing so, they challenge the generalized heterosexual normativity of the dominant confessional lineage and of contemporary poetry more broadly.

Rationale: The progressive slant here is both detectable and intense. The description doesn't merely study gender and sexuality — it uses language like 'radical self-disclosure,' 'trouble,' and 'resist...heteronormativity,' actively adopting the political posture of the writers it covers. This advocacy register is what pushes it from a 2 to a 1.

Example 2 (Rating: 2)

Course Description:

When thinking about an individual's health, we commonly begin with the physical body. Yet bodies are never separate from their surroundings: how people feel, whether illness occurs, and even lifespan are shaped by forces beyond biology. In this course, we will turn attention to the environments in which bodies live, considering how social conditions influence health, medical care, and well-being. We will explore the societal drivers of health outcomes, including residential contexts, interpersonal networks, healthcare institutions, inequities, and structures of power.

Rationale: A coder can detect a progressive slant here: the course centers inequality, power structures, and social determinants of health — topics associated with progressive intellectual traditions. The framing ('explore,' 'consider') is academic rather than activist, so the slant is present but not intense. This is a 2, not a 1.

Example 3 (Rating: 3)

Course Description:

policy evaluation: prepares students to systematically examine governmental responses to collective problems by specifying issues, developing alternative courses of action, identifying the stakes of decisions, and forecasting likely consequences. expects prior exposure to statistics, microeconomic reasoning, and the institutions and procedures of government.

Rationale: While clearly political in subject matter, the course is focused on the evaluation of policies broadly using analytic tools, rather than advocating for or engaging with topics primarily associated with one side.

Example 4 (Rating: 4)

Course Description:

An interdisciplinary study of competing and often mutually inconsistent modern arguments for the ethical legitimacy of capitalism, emphasizing market-oriented, objectivist, and christian forms of justification, approached through historical, economic, political, and philosophical perspectives. Readings include classical political economy, socialist critique, pro- and anti-slavery writings, reformist and interventionist accounts, conservative public policy, free-market theorists, monetarist economics, and religious defenses of enterprise...

Rationale: A coder can detect a conservative slant: the course is organized around defending capitalism's morality, concentrating on economic, objectivist, and Christian arguments — intellectual traditions associated with the right. It includes Marx and Keynes as counterpoints, and the tone is academic ('examination,' 'considered historically'), so the slant is present but not intense. This is a 4, not a 5.

Example 5 (Rating: 5)

Course Description:

Broad survey of political and economic thought, governing structures, market mechanisms, civic questions, fiscal policy, and public and commercial participation, stressing the integral connection between constitutionally restrained government and an entrepreneurial market order while presenting a christian perspective on the proper role of government and economics.

Rationale: The conservative slant here is both detectable and intense. The course doesn't merely study the relationship between government and markets — it emphasizes their inseparability and frames the entire subject through a 'christian worldview.' The link between limited government, free enterprise, and Christianity is treated as a premise, not a hypothesis. This advocacy register is what pushes it from a 4 to a 5.