Separation of Church and State Curricula? Examining Public and Religious Private School Textbooks*

Anjali Adukia† Emileigh Harrison‡

March 2026
Active Working Draft

Abstract

Curricula impart knowledge, instill values, and shape collective memory. Given the importance of what children are taught for their worldviews and later life outcomes, little is known about how the topics and values conveyed in educational materials differ across schooling environments in the United States. We examine 261 textbooks commonly used in Texas and California public schools, religious private schools, and home schools, applying natural language processing, computer vision, and generative AI tools to measure the presence and portrayal of people, topics, and values over time. Despite narratives of political polarization, we find few meaningful differences between public-school textbooks from Texas and California. By contrast, religious-school textbooks place substantially more emphasis on religious values and character values; contain less female representation; depict lighter-skinned pictured characters; and discuss topics such as religion and evolution differently. Important similarities also emerge: all collections rarely discuss LGBTQIA+ topics, and females are portrayed in less active and powerful contexts than males. Using survey data, we show that differences between religious and public-school textbooks align with differences in the beliefs and demographics of the populations these books serve. To explain the strong overlap between Texas and California textbooks, we develop and test a framework in which publishers serve a common public-school adoption market under state standards and costly customization. Consistent with this mechanism, we find that revisions to one state's content standards spill over into textbook content across states.

*For helpful feedback, we thank Ben Arold, Elliott Ash, Sam Bazzi, Marianne Bertrand, Dan Black, Ethan Bueno de Mesquita, Judy Chevalier, Michael Dinerstein, Steven Durlauf, Alex Eble, Martin Fiszbein, Sendhil Mullainathan, Elias Papaioannou, Heather Sarsons, Thomas Wollmann, David Yanagizawa-Drott, Hiro Yoshikawa, and seminar participants at AEFP, APPAM, Boston University, CESifo, Columbia, FGV, Harvard, Johns Hopkins, London Business School, RAND, Stanford, SREE, UChicago, UCLA, Univ. of Michigan, UPenn, Warwick, BFI AI in Social Science conference, Yale, and the Zurich AI/Economics conference. For exemplary research support, we especially thank Matt Bonci. We also thank Sam Abers, Elisa Xi Chen, Paula Dastres Gallardo, Benjamin Leiva, Jake Nicoll, Celia Zhu, and other members of the MiiE Lab for excellent research assistance. For financial support, we thank the National Academy of Education, Spencer Foundation, and Becker Friedman Institute at UChicago. The research reported here was also supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A200478 to the University of Chicago. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

† University of Chicago, MiiE Lab, IZA, and NBER, adukia@uchicago.edu

‡ University of Texas at Austin, MiiE Lab, and the University of Chicago Inclusive Economy Lab, emileigh.harrison@austin.utexas.edu

Our collective memory and understanding of the world are shaped by a variety of forces including the stories we share (Michalopoulos and Xue, 2021; Adukia et al., 2023), the narratives propagated through popular culture (Riley, 2022; Michalopoulos and Rauh, 2024), as well as the monuments we build and the history they represent (Bazzi et al., 2023). These elements not only influence *what* is remembered but also determine *who* is remembered and *how*. Education plays a pivotal role in this process with curricular materials serving as critical sites of learning. These materials do more than convey academic knowledge; they also instill children with societal values such as responsibility, kindness, and citizenship. Consequently, changes in educational content can shape children's worldviews, influencing their perspectives both explicitly and implicitly through the discussion of certain topics and values, who are represented, and how they are portrayed (Cantoni et al., 2017; Riley, 2022; Arold, 2024; Arold et al., 2025).

Parents recognize the role of education in transmitting values and seek out learning environments that reflect the particular sets of values, cultural norms, or religious beliefs they want to instill in their children (Catt and Rhinesmith, 2016). This behavior is consistent with theories of ideological threat and cultural transmission, which predict that parents choose institutions that reinforce and transmit preferred beliefs across generations (Wheaton, 2020; Bisin and Verdier, 2001). Evidence suggests that some parents are even willing to sacrifice academic quality in order to enroll their children in schools that more closely align with their values (Abdulkadiroğlu et al., 2018; Williamson and Withrow, 2025). The recent expansion of school choice programs has further enabled this behavior by allowing parents to direct public funds toward private or home-schooling options that they perceive to be a better values match. However, the minimal regulatory oversight governing the curricula of many private schools and homeschooling environments means that we know very little about the values these children are being taught outside the public school system. The substantial and increasing allocation of tax dollars to these private alternatives and the potential long-run consequences of educational content for children's beliefs and development motivate the need to understand what students are outside the public school system. This study addresses this gap by systematically examining the values embedded in the textbooks used in public, private, and home schools across different political and religious contexts.

We examine state-adopted public school textbooks in California and Texas, states that serve the largest student populations in the U.S., which allows us to understand the content being provided to a large portion of children in the U.S. and how educational materials vary across different political landscapes. We also examine curricular materials often adopted in religious private school and home school contexts: those produced by conservative Christian

publishers Abeka, Accelerated Christian Education (ACE), and Bob Jones University Press (BJU Press), which allows us to understand how educational materials in religious settings compare to public school materials. We focus our analysis on textbooks for the foundational subjects of Reading, Science, and Social Studies and those intended for third and fifth grades.

We apply computational social science methods to measure the representation of values, topics, and people across textbooks. Overall, our analysis reveals meaningful parallels between the public school collections, while our collection of religious school textbooks differ notably, featuring more character values, less female representation, characters with lighter skin, more White individuals, and differential portrayal of topics such as evolution and religion, including in science textbooks. Consistent with survey data on beliefs related to what values should be taught to children, religious textbooks emphasize obedience more than public school textbooks while public school textbooks place a greater emphasis on critical thinking. Religious topics and values are present in all three collections. In public school textbooks, however, they are largely limited to the context of the nation's founding, whereas in the collection of religious school textbooks, they are discussed in the context of theological concepts such as heaven or prayer. Important similarities, however, also emerge: for example, each collection portrays females in contexts that are more positive but less active and powerful than males, discusses Christianity more than other religions, rarely discusses historical or famous individuals who are Asian, Indigenous, or Latine, who were born outside of the North America and Europe, or who identify as lesbian, gay, or bisexual, and depicts the U.S. founding era and slavery in similar contexts.

Using survey data from the General Social Survey, we show that some of these differences and similarities between religious textbooks and state adopted textbooks reflect differences and similarities in beliefs and values between the populations these textbooks serve, suggesting that parents are choosing educational environments that reflect their values to preserve cultural transmission. We also show that the large overlap between Texas and California textbooks is consistent with the principle of minimal differentiation (Hotelling's law) and that differences in state content standards spill over across states.

Our study contributes to the literature focused on the content analysis of children's literature and educational materials (Wade, 1993; Allen et al., 1993; Clark, 2007; Koss, 2015; Braden and Rodriguez, 2016; Bickford and Knoechel, 2017; Crisp et al., 2018; Stevenson and Zlotnik, 2018; Koss et al., 2018; Koss and Paciga, 2020; Lucy et al., 2020). The innovative uses of computer vision and natural language processing tools we develop and apply throughout our analysis contribute to the growing body of computational social science literature in a manner similar to Caliskan et al. (2017); Garg et al. (2018); Kozlowski et al. (2019); Adukia

et al. (2022a,b, 2023); Voth and Yanagizawa-Drott (2024). For example, we introduce a novel approach to analyzing the contexts in which characters are depicted in images: we use generative AI to produce standardized descriptions of each character's setting and actions, and then apply word- and document-embedding methods to quantify how these portrayals differ across identity groups – capturing not only how often different groups appear, but the semantic, emotional, and domain-specific contexts in which they are shown. These methodological advancements can have meaningful implications for expanding the broader field of computational social science and the scope of research inquiry that can be pursued.

The paper proceeds as follows: Section I provides background. Section II discusses prior content analyses. Section III describes the textbooks in our sample. Section IV explains our image and text processing tools. Section V presents our findings. Section VI develops a framework for how adoption markets, publisher incentives, and state-level regulation jointly shape textbook content and Section VII concludes. The appendices contain additional findings and further information about the methods.

I Parental Values and the Expansion of School Choice

Knowledge and values are transmitted to children through two primary pathways: vertical transmission from parents and family, and horizontal transmission from society and peers. Vertical transmission occurs from the beginning of a child's life in their home, and public schools have traditionally served as a key vehicle for horizontal transmission, imparting not only academic knowledge but also broader societal values. A growing body of evidence shows that changes in public educational content and instruction have a causal impact not only on children's later life outcomes but also on their worldviews. Changes in high school politics curricula can affect students' views on democracy and free markets (Cantoni et al., 2017) and changes in public school content standards related to evolution can affect students' probability of choosing a STEM career (Arold, 2024), messages about the abilities of others with similar racial and gender identities can impact children's academic performance (Riley, 2022), and the removal of compulsory religious education can reduce the religiosity of students in adulthood (Arold et al., 2025). Given the importance of what we teach our children on their later life outcomes, debates about what should be taught – and what should be excluded – from public school curricula have been going on for centuries. In U.S. colonial times, schools were deeply intertwined with Christian instruction. However, as the nation formed, tensions between religious influence in public institutions and the ideal of a secular state led figures like Thomas Jefferson to advocate for a "wall of separation between church and state." This principle has continued to shape legal battles in education, including debates over school-sponsored prayer (e.g. *Engel v. Vitale*) and the teaching of

evolution (e.g. *Epperson v. Arkansas*, *Tennessee v. Scopes*). More recently, curricular debates have expanded beyond religion to include questions of how history and identity are taught in schools. Some states, like California, have embraced inclusive standards, such as the 2011 FAIR Education Act, which mandates broader representation of ethnic, gender, LGBTQIA+ groups, and persons with disabilities in history instruction. By contrast, other states have introduced "educational gag orders," seeking to restrict discussions of race, gender, and American history (Friedman and Tager, 2022). Notably, these standards only apply to public schools and not private schools or homeschooling families.

Parents recognize the role of education in transmitting values and seek out learning environments that reflect the particular sets of knowledge, cultural norms, or values they want to instill. Faced with either dissatisfaction over academic standards or misalignment between their values and values prioritized in public education, some are increasingly turning to alternative school choice options, including private and home schools through the use of vouchers (EdChoice, 2023). This behavior is consistent with theories of ideological threat and models of cultural transmission, in which individuals respond to perceived ideological threats by selecting institutions that reinforce preferred beliefs across generations (Bisin and Verdier, 2001; Tabellini, 2008; Wheaton, 2020). However, evidence on the academic benefits of school choice programs is mixed. Some research has found that academic improvements can occur, but only when students gain access to higher-quality schools (Deming et al., 2014), but Abdulkadiroğlu et al. (2018) find that many parents in rural settings use private school vouchers to move their children to lower quality schools as measured by students' academic performance. These findings suggest that some parents may be motivated by factors beyond academics, such as the desire for educational environments that are better aligned with their values. Survey data support this hypothesis: in a survey of parents with children in private schools in Indiana, more than 80% listed religious, moral, or character values instruction as a top desired school quality compared to approximately 65% who listed better academics (Catt and Rhinesmith, 2016). A desire for religious values instruction is reflected in private school enrollment patterns: most students educated outside the public school system are educated in a religious school environment.[1] Prioritizing social values over academics in school choice is not a new phenomenon. In the 1960s and 1970s, some parents reacted to the desegregation of public schools by organizing all white private schools, which had a negative impact of the

[1]In 2017, among the 5.8 million children who were enrolled in a private school, 76% were in a religious school, with 49% enrolled in a conservative Christian school (National Center for Education Statistics, 2019). 75% of parents who chose to educate their children at home listed a desire to provide moral instruction as a top reason why they choose to homeschool (National Center for Education Statistics, 2022). Pre-COVID estimates of the number of homeschoolers who identify as being religious, or for whom religion is a primary motivation for homeschooling, ranged from over 50% to 90% (Kunzman, 2010; Bartholet, 2020).

academic outcomes of white students exposed to a segregation academy, and had no effect on Black students (Williamson and Withrow, 2025). Taken together, these findings indicate that many parents are willing to sacrifice educational quality for academic environments or curriculum that better align with their values.

Public policy has rapidly evolved to accommodate parents' demand for more educational options. School choice programs such as vouchers, tax-credit scholarships, and education savings accounts have proliferated in the United States, especially in the past few years. These programs make it easier than ever for parents to direct public education funds toward private and home schools that reflect their preferences. Recent U.S. Supreme Court decisions (e.g. *Espinoza v. Montana Department of Revenue* and *Carson v. Makin*) have also struck down barriers to using public funds at religious schools, ruling that once a state subsidizes private education, it cannot exclude faith-based schools from participation. Additionally, Congress recently authorized the nation's first federal tax-credit scholarship program, signaling bipartisan momentum to expand parental choice at an unprecedented scale which means parents today have more opportunities than ever to opt out of the public school system and still obtain public funding for the private school environments they perceive as better aligned with their values. However, this movement also raises pressing questions about what is being taught in these alternative settings and how it compares to the content of public education.

Despite the growing public investment in private and home schooling, there is surprisingly little oversight or transparency regarding the curricula and values being taught outside the public school system. Private schools and homeschooling families face minimal or no curricular oversight.[2] Consequently, we lack basic knowledge of the curricula and values being transmitted outside the public school system. This lack of information makes it difficult to evaluate the broader societal impacts of diverting public resources to private education. To ensure that school choice policies serve the public interest, it is crucial to understand the content of what different schools teach. By empirically examining the values embedded in public, private, and home school curricula, we aim to shed light on this underexplored dimension of education. A clearer picture of the values education landscape can help guide policy decisions about how to allocate resources and design oversight in a way that respects parental choice while promoting the shared goals of our society.

[2]Most states require private school curricula to be "similar" to public standards, but 18 states either have no requirements or only require that private schools teach students the state constitution and state history. While states often have similar curricular requirements on paper for homeschoolers as they do for private schools, there is rarely any enforcement for homeschooling families (Bartholet, 2020).

II Prior Research

Our study complements existing qualitative and quantitative work examining the content of curriculum in the United States and around the world. Much of this work focuses on the representation of gender and race in educational materials. For example, Crawfurd et al. (2024) analyze 1,255 English-language textbooks from 34 countries and find that women tend to be portrayed in stereotypical gendered roles. Lucy et al. (2020) also uses NLP tools to examine who is present in the text of a sample of 15 Texas high school textbooks used between 2015 and 2017 and find that Latine people are rarely discussed, the most common famous figures are nearly all White men, and that women tend to be discussed in the contexts of work and the home. Another related paper by Stevenson and Zlotnik (2018), focused on college-level introductory economics textbooks most commonly purchased in the U.S., finds that fictionalized women are more likely to be involved in food, fashion, or household tasks.

A smaller body of research examines values transmitted through textbooks (c.f. Sallabaş 2013 who conducted a qualitative analysis of character values in Turkish textbooks). Even fewer studies examine textbooks commonly used in religious settings. The most relevant is a qualitative analysis by Scaramanga and Reiss (2018) on curricula published by Accelerated Christian Education (ACE) – one of the textbook publishers we include in our study – which found that churches and schools are often depicted as racially segregated in their materials.

Finally, most of the previous work discussed above focuses solely on the text but largely ignores the images shown in textbooks. One exception is Deckman et al. (2018), who manually examined 1,468 images from U.S. elementary and middle school health textbooks to study the portrayal of race, gender, and sexuality. Despite finding substantial *presence* of gender and racial diversity, they saw that women and people of color were often portrayed in stereotypical roles, and no depictions of same-sex couples appeared in any of the images. In particular, we present a novel application of generative AI and word embedding methods which allow us to generate descriptions of the context in which a character is pictured and use traditional word embedding methods to measure the portrayal within images. This provides us insight not only into how often characters of different identities are mentioned but also the contexts in which they are pictured

Our study expands on the textbook analysis literature by applying computational content analysis to both the images and the text of 261 primary school textbooks used in public and religious educational settings from 1980 to the present. Previous studies have typically limited their analyses to cross-sectional examinations of textbooks in a single

setting. Our panel of textbooks affords a unique opportunity for longitudinal analyses across different educational settings and over multiple decades. Our broader set of tools enables new analyses, such as examining the representation of skin color in textbook images, an important but often overlooked dimension of human categorization (Keith and Monroe, 2016). Finally, we systematically measure the presence and portrayal of people, topics, and values over time in a single study. Our approach enables us to replicate qualitatively similar results found in previous work while also documenting new patterns. Our methods advance both image and text analysis in the social science literature along with recent work by Adukia et al. (2022a,b, 2023) examining children's books, Ash et al. (2021) examining newspapers, Voth and Yanagizawa-Drott (2024) examining yearbooks, and Kamel and Woo-Mora (2023) examining sports statistics.

III Data Sources

In this study, we focus on curriculum as *written*, and specifically, the curriculum as it is presented through textbooks. We draw data from three distinct sets of elementary school textbooks used in: (1) Texas public schools, (2) California public schools, and (3) religious private and home school settings. We digitize a novel data set of Reading, Science, and Social Studies textbooks written for the third and fifth grades. Our data set comprises over 75 thousand pages originating from 261 different textbooks, amounting to a corpus of approximately 16 million words and 60 thousand detected faces (Table 1). [3]

III.A Texas State-Adopted Textbooks

We study Texas public school textbooks for four primary reasons. First, Texas adopts textbooks at the state level which means we can construct a panel of textbooks that children were likely to have been exposed to over time. Second, Texas serves one of the largest student populations in the U.S., so the messages that are sent in these books are likely to reach a broad audience. Third, because they are such a large share of the textbook publishing market, Texas has historically influenced the content of textbooks used across the United States. Finally, Texas is often portrayed as reflecting conservative values, both politically and in the media, which provides a view into how one side of the ideological spectrum presents education in the public sphere.

The state of Texas, through its State Board of Education, adopts instructional mate-

[3]Note that our sample is not perfectly balanced across subjects, grades, and year in use across collections. Because of this, when comparing average measures across collections, we make sure to first calculate the measure for each collection, subject, grade, and year in use and then average these measures within a collection, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages. This approach allows us to attribute observed differences across collections to actual differences in content, rather than to shifts in the composition of the sample over time.

rials and maintains an annual "state-adopted" list (the Current Adoption Bulletin) for K–12 grades. Prior to 2011, instructional materials funding was largely tied to this list, and districts typically ordered textbooks through the state system, with the state paying publishers directly. Consequently, the adopted list provides a reasonable proxy for the textbooks in use. In 2011, Texas shifted to a more flexible funding regime (the Instructional Materials Allotment), allowing districts to purchase state-adopted materials or other instructional materials, subject to certification that they cover the Texas Essential Knowledge and Skills (TEKS). In light of this, we first digitized a convenience sample of approximately 50% of all textbooks on the yearly state-adopted textbook list in use between 1985-2010. Then, for the years 2011-2021, we focused our book digitization on the top two textbooks purchased by school districts each year for each relevant grade and subject (reliable textbook purchasing data were not collected before 2011).

III.B California State-Adopted Textbooks

Similar to Texas, California also adopts elementary school textbooks at the state level and has a large influence on the textbook publishing market due to the size of its student population. It is typically associated with more progressive policies and is often posed as the political or substantive converse of Texas. Having these two political extremes in our sample will allow us to bound the most conservative and the most liberal implementations of traditional public school curriculum.

California curates a list of state-adopted textbooks for grades K-8 and updates it every couple of years for each subject.[4] We digitize the universe of books in the relevant grades and subjects which were adopted in California for all years in which the list of adopted materials are available. This includes adopted textbooks beginning in 1999 for Social Studies, 2000 for Science, and 2002 for Reading and ending with the books that were in use as of 2022.

III.C Textbooks by Religious (Conservative Christian) Publishers

A third corpus that we examine comprises religious school textbooks published by three ideologically conservative Christian textbook companies: Abeka, Bob Jones University Press (BJU Press), and Accelerated Christian Education (ACE). We refer to this set of books as our religious collection. These textbooks are among the most commonly used by homeschooling families (Kunzman and Gaither, 2020) and adopted by over a third of non-Catholic Christian schools receiving taxpayer funding (Klein, 2017). We focus on textbooks published by these three Christian publishers both because of their wide adoption, but

[4]Note that during the economic crisis of 2007, California did not make any changes to their state-adopted textbook lists to save districts the expense of purchasing new textbooks.

also because there are many taxpayer dollars funding schools which use these curricula.[5] We digitize the universe of textbooks published by these three publishers for the grades and subjects in our sample starting with books published in 1980 up until 2022, including updated editions of the same textbooks.[6]

Using Google Trends search data as a measure of interest in these publishers (Figure A.1), we see that *Abeka* was the most commonly searched publisher, but after the onset of the COVID-19 lockdowns in March 2020, there was a marked increase in searches for both *Abeka* and *BJU Press*. We also see that there was a steady decline in search interest for *homeschool* over time, but then it spiked once schools shut down and then began to track the search interest for *Abeka*. This is consistent with evidence that since the onset of COVID, the number of parents choosing to homeschool their children has increased substantially (Musaddiq et al., 2022; Jamison et al., 2023). Notably, the COVID-era increase in homeschooling was likely not driven primarily by religious motives, but we still see increases in religious homeschooling curriculum. This is likely because there are few non-religious comprehensive curricular options that families can easily adopt when they choose to homeschool their children (Kaleem, 2016).

IV Methods

In order to analyze large corpora of educational materials, we leverage tools from the computational social sciences, including artificial intelligence methods, which allow for a cost-effective and systematic approach to content analysis. We treat each scanned textbook page as both (i) an image and (ii) a source of printed text. The process that we use to convert images into data is illustrated in Figure 1a, and the process that we use to convert text into data is presented in Figure 1b. Additional details for all methods including, complete prompt specifications, pre-processing steps, and regression details are provided in Appendix D.

IV.A Images As Data

IV.A.1 *Character Detection, Face Detection, and Feature Classification*

One way that we extract data from images is by using a Large Multimodal Model (LMM), which enables large-scale analysis of visual content, including the detection of pictured characters, their characteristics, setting, and actions. Importantly, unlike typical face detection models, this allows us to detect characters whose faces are not visible, as well

[5]Notably, in Indiana, approximately 4,240 students received over $16 million in scholarships to attend schools that use the Abeka or BJU Press curricula, according to 2016-17 figures from the Indiana Department of Education (Klein, 2017).

[6]Note that most of the curriculum published by ACE consists of workbooks called PACEs (self-guided Packets of Accelerated Christian Education) instead of textbooks, so we only include books from their Reading curriculum in our analysis.

as a larger number of non-humanoid characters. Using Gemini 2.5 Pro Preview, we extract two types of data from each textbook page containing an image: scene-level metadata which captures contextual information including setting descriptions and depicted activities and character-level data which contains demographic attributes including race, gender, and human versus non-human classification for each pictured character we detect.

We supplement this character detection with a CNN-based face classifier deployed on Google's Vertex AI, trained to detect faces in both photographs and illustrations (Szasz et al., 2022; Adukia et al., 2023). Our main results rely on race and gender predictions for characters from our LMM. However, we also predict the race and gender of each detected face using a CNN feature classification model trained on a manually curated data set of both illustrations and photographs (Szasz et al., 2022). In the sample of detected characters that can be matched to a detected face, the race and gender predictions of the LMM outperformed this CNN model; validation exercises are discussed in Appendix C.

IV.A.2 *Skin Color*

To measure skin color, we segment facial skin pixels and construct a continuous measure of skin tint based on the L* value of a representative skin color in CIE L*a*b* space (0 = darkest; 100 = lightest), following Adukia et al. (2023).

IV.A.3 *Image Description Embeddings*

To compare visual portrayals at scale, we construct embeddings from the LMM-generated text describing each pictured human character. For each character, we concatenate the model's descriptions of actions, location, and what is happening into a single short document and estimate a 300-dimensional Doc2Vec model on these character-documents. We use these embeddings in three ways: (i) we construct a textbook-level embedding by averaging the (normalized) character-description embeddings within each textbook and compare these book-level embeddings across collections; (ii) we identify the words most distinctively associated with descriptions of female versus male pictured contexts and examine the valence, arousal, and dominance of those words; and (iii) we measure the similarity between image descriptions and pre-specified semantic domains such as family, sports, home, and politics.

IV.B Text As Data

IV.B.1 *Large Language Models*

After extracting text from the scanned textbook pages using Google Cloud Vision API's OCR, we measure the topics and values included in the text using a large language model (LLM). We use OpenAI's GPT-4 model to analyze single pages of textbooks and identify topics and values present in them. We validate the model's topic classification by

counting how often topic-related words appear on pages that were (or were not) classified as containing that topic by GPT (Figure C.1). Additional prompt details and validation exercises are provided in Appendix D and C. We use GPT to classify topics present on each page; this helps us generate hypotheses about unanticipated topics and systematically test for differences in topic prevalence across collections (Ludwig and Mullainathan, 2024).

IV.B.2 *Token Counts and Named Entity Recognition*

We use token counts to measure the prevalence of gendered terms and specific topics. We also count named entities including locations, events, and people using Named Entity Recognition (NER). To measure geographic exposure, we use the "GPE" and "LOC" tags; to identify historical events, we use the "EVENT" tag. To identify famous individuals, we merge entities tagged as "PERSON" with the Pantheon 2.0 dataset (Yu et al., 2016). We use remaining "PERSON" entities that are not matched to Pantheon to measure mentions of non-famous individuals by extracting first names and estimating the probability that each name is female or male using Social Security Administration name-frequency data.

IV.B.3 *Word, Page, and Textbook Embeddings*

We use embeddings to characterize differences in semantic contexts across people, topics, and textbooks. For text, we construct pronoun-instance à la carte (ALC) word embeddings using the `conText` framework and use these embeddings to (i) identify top relative words by gender and their VAD scores and (ii) measure gender skew in similarity to pre-specified domains. We also estimate Doc2Vec embeddings at the page and textbook level and integrate GPT topic labels with page embeddings to construct topic embeddings. We use embedding regressions to formally test for differences across collections. Regression specifications are provided in Appendix D.

IV.B.4 *Sentiment (VAD Lexicon)*

To characterize sentiment, we use the NRC VAD Lexicon, which provides valence, arousal, and dominance scores for over 20,000 English words. Valence (V) measures positivity (happy to unhappy), arousal (A) measures affective activation (active to passive), and dominance (D) reflects emotional control (powerful to submissive).[7]

V Results

In this section, we examine the content patterns across our textbook collections, beginning with the *overall similarity* between collections by subject and over time.

[7]For example, "teamwork" and "leader" have V, A, D scores of 0.92, 0.49, 0.77 and 0.83, 0.58, 0.93, respectively – both positive in valence, but "leader" scores higher on arousal and dominance.

V.A Overall Patterns in Textbooks

First, we analyze the overall similarity of textbook content (both text and images) using document embeddings to cosine similarity between textbooks in embedding space. Higher cosine similarity between textbooks indicates more similar semantic content in both the text and the described settings and actions of pictured characters.

We begin with a descriptive view of the embedding space. Figures 2a and 2b plot book-level text and image embeddings (based on image descriptions) in two dimensions using t-SNE, with colors denoting collections and shapes denoting subjects.[8] Two patterns stand out. First, books cluster strongly by subject, suggesting that the embeddings recover meaningful differences in curricular content. Second, within subject, California and Texas textbooks tend to occupy similar regions of the space, whereas the religious collection is more clearly separated from both.

We formally test these visual impressions using normed coefficients from book-level embedding regressions in the spirit of Rodriguez et al. (2023) in Figures 2c and 2d, which provide measures of the overall magnitude of semantic difference in embedding space between groups, aggregated across all embedding dimensions. These regressions show that subject differences are large in both text and image embedding space. Differences between collections are smaller than differences between subjects and show that while California and Texas textbooks are not identical, they are substantially closer to one another than either is to the religious collection.

Figure A.3 reaches the same conclusion by calculating cosine similarities for textbook pairs from different collections that share the same subject, grade, and year in use. The histogram in Figure A.3a shows that some textbook pairs have near-perfect similarity scores, consistent with publishers reusing closely related content across state editions (See Section VI for a discussion of the textbook publishing process). Averaging similarities over time yields the same ranking in every decade: California and Texas textbooks are consistently more similar to each other than either is to the religious collection (Figure A.3b). Over time, however, Texas Science textbooks have become less similar to California and more similar to Religious textbooks, while California and Texas Social Studies textbooks have grown more similar to each other.

The next sections move from these aggregate embedding patterns to specific dimensions of content, allowing us to identify where California and Texas textbooks still differ and

[8]For visualization, we project the 300-dimensional embeddings into two dimensions using t-SNE; the figures should be space.

where the religious collection diverges most sharply from both.

V.B Values

We start with a discussion of the values present in each textbook collection. The top values are similar across collections, with the three most common values centered on learning. We see this overlap especially between California and Texas textbooks (Figure A.4). The primary outlier is that of *faith*, which shows up as a top value in the religious collection; this is consistent with the large proportion of religious values found in the religious collection which we discuss in more detail below.

Character Values Character values are an important part of early education. Texas codifies character education in state law, requiring instruction in four broad domains: trustworthiness, responsibility, caring, and citizenship (Texas State Board of Education, 2020). In contrast, California provides non-binding guidance, recommending a six-part framework: trustworthiness, respect, responsibility, fairness, caring, and citizenship (California Department of Education, n.d.).[9] Despite the importance that public schools place on character values through their standards, the majority of parents who choose private or home schooling options report "morals," "values," or "character" education as among the top most important qualities when choosing a school, suggesting that they assume their children may not get these through the public system (Catt and Rhinesmith, 2016).

In this section we examine whether religious collection has a larger focus on character values than public school textbooks, and whether there are differences between the proportion of character values in Texas and California that reflect differences in the strictness of the curriculum standards.[10] Figure 3a shows that the religious collection contains substantially more character values in Reading textbooks (70.8%) than either California or Texas (58.7%-55.4%), and a slightly higher (though not statistically significant) share of pages with character values in Science and Social Studies textbooks. Character values are nonetheless common in public school textbooks, appearing on roughly 55–59% of Reading pages, about 17–18% of Science pages, and about 40–45% of Social Studies pages. We find no statistically significant differences between Texas and California in the prevalence of char-

[9] These *Six Pillars of Character* align with national models like CHARACTER COUNTS! and mirror the four Texas pillars, though Texas bundles fairness and respect into citizenship.

[10] We focus on the four character value pillars that are explicitly mentioned in the Texas standards and that are mentioned in California guidance (we bundled respect and fairness into citizenship the same way that Texas does). Publishers of textbooks from our religious collection do not have a pre-defined list of character traits but Abeka's K5 development materials include additional traits not mentioned by Texas or California such as obedience and self-control, tolerance, and modesty. To incorporate these, we include two additional categories; an "Attitude" category that incorporates character values related to attitude or temperament such as cheerful or joyful and an "Other" category that includes all other un-categorized character values. A list of all the values included in the groupings can be found in Table B.1.

acter value categories outlined in Texas state standards and California's suggested guidance, despite differences in whether these values are mandatory (Figure A.5a). Similarly, when examining the most frequently mentioned character values across collections, we observe no meaningful differences between Texas and California. While the religious collection contains more character values overall, some individual values appear more frequently in public school textbooks (Figure A.5b); for example, creativity is mentioned on a larger share of pages in public school textbooks (11%–13%) than in the religious collection (5%).

Religious Values Religion is one of the most frequently cited reasons why some parents opt for private or home schooling, underscoring the weight placed on its value Catt and Rhinesmith (2016); National Center for Education Statistics (2022); Bartholet (2020). Given this and the fact that textbooks in the religious collection are published by Christian publishers, we would expect these textbooks to contain a larger proportion of religious values. We test the extent to which this is true by combining values related to religion such as *religious faith, religious freedom, religious beliefs, religious diversity, religiosity, religious tolerance*, and *religious devotion.* Using this group of values, we find that religious values appear substantially more often in textbooks from the religious collection across all subjects (Figure 3b). In Reading textbooks, religious values appear on about 4.0% of pages in the religious collection compared to 0.4% in California and 0.2% in Texas. In Science textbooks, religious values are present on roughly 2.1% of pages in the religious collection and are essentially absent in both California and Texas textbooks. The largest differences appear in Social Studies textbooks, where religious values occur on 15.9% of pages in the religious collection compared to 5.6% in California and 2.9% in Texas. Together, these patterns show that while religious values appear occasionally in public school textbooks, they are far more prevalent in textbooks from the religious collection, particularly in Social Studies. In the next section, we show a similar prevalence of religious topics in our religious collection of textbooks.

V.C Topics

While the topics that are covered in textbooks are not explicit values, which topics are discussed, how often they are discussed, and the context in which they are discussed reflect the values of society. In this section, we document the topics emphasized in each of our textbook collections.

We first show the overall top topics that are addressed by subject (Figure A.4). We see that similar topics are discussed across collections with some notable exceptions unique to each collection. In Reading textbooks, many of the top topics (e.g. *family, education, nature, friendship,* and *adventure*) are the same across all collections. However, only the

religious collection includes *prayer* and *faith* among its top topics, while only public school textbooks include *reading comprehension*. In Science textbooks, *astronomy, biology, physics, observation*, and *science* are among the most common topics across all three collections (Figure A.4e). In Social Studies textbooks, *geography, exploration, history* (including *American history*), and *agriculture* are among the most common topics (Figure A.4f). Notably, *religion* appears as a top topic in Social Studies textbooks in the religious collection but not in public school textbooks. In addition, *slavery* appears as a common topic in California textbooks, while the *Civil War* is more prominent in the religious collection; neither appears among the top topics in Texas textbooks.

These overall patterns in the top topics indicate meaningful overlap in the topics most commonly taught to children, suggesting that children encounter similar material across different educational settings. In the next section, we focus not on the most common topics, but on a set of contentious topics.

High-Profile Topics Much of the public debate, does not revolve around the set widely shared topics discuss above. Instead, public discourse often focuses on more contentious issues such as evolution, slavery, racism, sexuality, and climate change.

We first examine the prevalence of these high profile topics as identified by a large language model, both overall and over time (Figures 4a and A.6). These figures show that California and Texas textbooks are much more likely to address topics related to climate change than textbooks in our religious collection. By contrast, our religious collection is the only set of textbooks that discusses creationism and is much more likely to address topics related to evolution (typically as a way of contrasting it to creationism). California textbooks are the most likely to include slavery-related topics overall, though patterns are mixed over time; whereas the religious collection is more likely to include topics related to the Civil War, and increasingly so over time. Both California and Texas discuss immigration more often than religious school textbooks, though in Texas textbooks we see an increase in the proportion of pages which discuss immigration over our sample period and in California we see a decrease. Racism and LGBTQIA+ related topics are rarely, if ever, mentioned across all three collections.

We have described the *prevalence* of certain topics across textbook collections. However, just as important as *whether* certain topics are discussed is *how* they are discussed. By integrating the information on topics with our embeddings as described in Section IV, we can examine differences in the context or sub-topics surrounding a given topic across collections.

First, we examine overall semantic differences between these topics across collections using topic embeddings regressions, following our approach shown in Figure 2c. For some topics we also plot topic embeddings for each collection along with the top word embeddings closest to each topic embedding in two-dimensional space to visually present and examine the context surrounding each topic embedding in Appendix E. We find that for most topics, discussions in Texas textbooks are significantly more similar to those in California textbooks than to those in our religious school textbook collection. For topics related to Christianity and the Civil War, differences between Texas and California are not statistically significant at the 5% level, for the other topics differences are significant but relatively small. Taken together, these results indicate that public school textbooks in Texas and California discuss most topics more similarly to each other than either does relative to religious school textbooks. Evolution is the main exception, with distinct patterns of discussion in all three collections.

We further explore differences in the way that Evolution is discussed across collections by examining Figure E.4 which shows the topic of evolution is often discussed in the context of words such as creation, theories, evidence, and Genesis in our collection of religious school textbooks but this is not the case in public school textbooks. Unlike many of our other topics which appear on hundreds to thousands of pages, the topic of evolution only appears on fewer than 100 pages in each collection (Figure 4a). We manually examined these pages and extracted representative passages to illustrate qualitative similarities and differences in the treatment of evolution across collections.

Pages in our collection of religious school textbooks that contain the topic of evolution fall under three main categories: (1) concepts related to evolution are discussed, but evolution is not named or mentioned (for example, from a 3rd grade Science religious school textbook classified as discussing natural selection: "A twig caterpillar is the same color as a twig. In fact it looks just like a twig. Birds which eat insects ... think the caterpillar is part of the tree."); (2) evolution is explicitly brought up and depicted as not being true; and (3) evolution is discussed in the context of Social Darwinism: "The idea that only the fittest people can survive in society is known as Social Darwinism. Because people are naturally sinful and selfish this belief was dangerous." Examples of (2) are common. One such passage discusses the human and dinosaur footprint found next to each other in Texas as evidence against evolution. A different example passage from a religious school 5th grade Science textbook reads as follows:

> "Evolutionists interpret the fact that dogs and cats are similar as being an evidence for evolution. Creationists on the other hand explain that God created living things with

> similar structures because they would need the structures to do similar jobs. Cats and dogs for example have similar teeth because they enjoy similar eating habits. If a job is to be done with greatest efficiency there is usually a best method for doing the job. God used the best most efficient design in his creatures so they would be well suited for living on the earth."

In our collections of public school textbooks, pages predicted to contain the topic of evolution discuss a wide variety of concepts including Neanderthals, fossils, natural selection, prehistoric animals, the Ice Age, and behavioral or structural adaptations. One difference between Texas and California textbooks is that in the Texas collection, even though there are discussions of natural selection, adaptation, and the Earth being millions of years old, there is never any explicit mention of the Big Bang theory or Charles Darwin. An example of a passage from a Texas 3rd grade Reading textbook reads as follows:

> "If eohippus does not look like a horse of today, how do we know that eohippus was a horse? How do we know that eohippus was not a member of the dog family? We know that eohippus is a member of the horse family because we have found skeletons of other horses that lived long ago. When we put the skeletons in a row we can see small changes. Picture 2 shows just how horses changed over millions of years. It shows that eohippus did not change into a dog. It changed into the horses of today."

In our California collection, not only are there explicit mentions of the Big Bang Theory and Charles Darwin, there is also one unique passage in a California 5th grade Social Science textbook called "Stories of How the World Began" which describes creation stories from different cultures. An excerpt from that passage reads:

> "How did the world begin? Scientists say there was a big bang and dust and particles of matter came together to form stars and planets. In the Western nations the Bible told of Adam and Eve and how God created the world in seven days. The Popol Vuh, the sacred book of the Mayan Indians, told this tale of creation..."

The results in this section show that, for the grades and subjects we study, the topics at the center of many curriculum debates are mentioned only rarely This suggests that parents, teachers, and policymakers must care not only about the intensive margin of how much space textbooks devote to these issues but also about the extensive margin of whether topics such as evolution, climate change, and LGBTQIA+ individuals appear at all. Given the importance of this extensive margin, in the next section we examine other low-frequency topics that are distinctive to particular textbook collections.

Exclusive Topics Despite the fact that we document many overall similarities across public school textbooks, there are still dimensions on which textbooks from Texas and California differ. In addition to the topics which are commonly discussed across collections, it is useful to understand what topics are exclusively identified in specific collections and not in other

collections (Figure A.8). We see that only California textbooks are identified as having topics related to California history and geography, including discussion related to the *Kumeyaay tribe*[11] and *California national parks*. Only Texas includes topics related to the practice and knowledge of *Texas Essential Knowledge and Skills (TEKS) state standards*, some places in Texas such as *Fort Worth*, in addition to topics such as *civic leadership, the loch ness monster, the Berlin Airlift, the Special Olympics, etc.* Only the religious school textbooks have topics related to *God, evangelism, and creationism*. We further explore the extent to which textbooks in the religious collection discuss religious topics.

Religious Topics Consistent with the patterns for religious values discussed in Section V.B, Figure A.7 shows that religious topics appear on a much larger share of pages in the religious collection (12%, compared with 4.4% in California and 2.8% in Texas). Moreover, when a specific religion is mentioned, it is overwhelmingly Christianity. These differences in prevalence do not, however, tell us whether religion is discussed in the same way across collections. To examine differences in framing, we turn to the topic embeddings. Figure E.6 shows that in the Religious collection, the *Religious* topic embedding lies closer to words associated with God and Christianity, such as *savior, pray, creator, sins,* and *heaven*. In the California and Texas collections, by contrast, it lies closer to words associated with other religions and the U.S. founding era, including *Quaker, Islam, Amish, liberties, tolerance, colony,* and *settlers*.

Geography and History Political and media discussions of curricula are commonly centered around which historical figures and events should be highlighted and how they should be portrayed (Friedman and Tager, 2022). This discourse reflects concerns expressed by parents as highlighted in a survey conducted by the Heritage Foundation which found that 59.5% of teachers and 56% of parents surveyed noted the importance of schools including content related to the Constitution and the founding era of the United States (Sailor et al., 2021). As we see in Figure A.4, geography and history were among the top Social Studies topics identified in each collection. When examining specific mentions of places (Figure 5a), we see that children are primarily exposed to mentions of the Earth generally and the United States overall with very little relative exposure to other parts of the world. California is most likely to include mentions of California-specific geography, and Texas is most likely to include mentions of Texas-specific geography. The top mentioned places outside of the United States are either in Europe or in outer space.

Figure 5b shows the most commonly mentioned historical events. Most of these events occurred in the United States, while the remaining events are wars in which the United States was involved, namely World War I and World War II. Although the overall set of events is

[11] an Indigenous people from what is now Southern California

similar across collections, the frequency with which they are discussed varies. In particular, mentions of the Civil War increase over time in the California and Religious collections but decline sharply in Texas textbooks over the past two decades.

In addition to historical events, we also examine the representation of historical figures and prominent individuals. History and science came from individuals, so in the conveyance of information, textbooks often include discussion of these individuals. They can serve as real-life characters in our histories, examples of leadership, and as role models to whom children may aspire to be similar. The most frequently mentioned famous people in each collection are shown in Table B.2. George Washington, Thomas Jefferson, and Benjamin Franklin – who are all Founding Fathers – are among the most commonly included famous individuals across all collections. When we examine the birthplaces of the famous individuals to whom children are exposed (Figures 5c and A.11), we see relatively similar patterns over time and overall across public school textbooks, with most inclusion of individuals born in North America, followed by individuals born in Europe. Most of these individuals are male, but when females are present, they are mostly from North America. The religious collection has similar inclusion of individuals born in North America and Europe. For all of the collections, there is little to no representation of individuals from Africa, Asia (including East Asia and South Asia), the Middle East, and South America. The religious collection only has White males in the top twenty list of historical figures included. The most frequently mentioned individuals in each collection are all identified as being heterosexual, with the one exception of Jane Addams, a civic leader who identified as a lesbian, mentioned in the Texas collection (though her sexuality is never mentioned in the books themselves). In the next section, we further discuss the race and gender of famous individuals mentioned within textbooks.

V.D People: Representation of Race, Skin Color, Sexuality, and Gender

The people who are highlighted in the stories and history presented to children reflect societal views about whose stories are told and the roles they are expected to occupy in society. In this section, we discuss the representation of race, skin color, sexuality, and gender of people who are portrayed to children through images and through the discussion of historical figures.

Race. We measure racial representation separately in text and images. In the text, we cannot reliably infer the race of fictional characters from names alone, so we restrict our analysis to historical and other prominent individuals (hereafter called famous individuals) identified using Named Entity Recognition. In images, we measure the race of all detected human characters using the LMM output.

We find that the vast majority of famous individuals mentioned are White, accounting for 83% to 92% of famous individuals across collections and subjects (Table 1). This pattern is also visible in images and remains stable over time (Figure 6). Relative to public school textbooks, textbooks in the religious collection include more White and fewer Black famous individuals and pictured characters. Across all collections, Asian, Latine, and Indigenous individuals are rarely represented in either text or images, especially relative to the share of Latine students in California and Texas (Figure 9). California public school textbooks contain the highest shares of pictured Asian and Latine characters, though these remain modest at about 10–14%, compared with 7–11% in Texas and 6% in the religious collection (Table 1). Within racial groups, males are often more likely to be represented than females in both text and images (Figure A.10).

Skin Color. Skin color is one of the first ways that children learn to distinguish humans, and is distinct from (though highly related to) measures of putative race (Kessen and Weiskopf, 1976; Katz and Kofkin, 1997; Kinzler and Spelke, 2011; Adukia et al., 2025). In Figure 6c, we present the distribution of detected faces according to skin tint, measured on a scale from 0 to 100 representing how dark or light their skin color appears. We plot these distributions over a scatterplot showing the skin color of every face in our sample. Figure 6d shows that the patterns are relatively stable over time, with a slight trend towards increased lightening of character skin color in the religious collection over time.

Overall, we see that the pictured individuals in the public school collections from Texas and California are more likely to be depicted with darker skin on average than the pictured individuals in the religious collection. This pattern holds even when conditioning on race; in the religious collection the skin color of pictured faces is lighter than pictured faces in California and Texas textbooks, both on average and over time (Table B.3). This is true even when we condition on the predicted race of the face. We also see that older children (children in grade 5 as opposed to grade 3) are more likely to be shown characters with darker skin.

Sexuality. We also examined the number of mentions of famous individuals who identified as Lesbian, Gay, or Bisexual and found fewer than 20 individuals mentioned in each of our public school textbook collections and fewer than 10 individuals mentioned within our religious collection of textbooks. In Table B.5, we list each of these individuals and the number of times they are mentioned and the number of books they are mentioned in. In all three collections, the total mentions of LGB famous individuals comprise fewer than 1% of all mentions. Note that only Harvey Milk was mentioned as identifying as gay, while none of the other famous individuals were discussed in the context of their sexuality.

There were only two books in our sample that we identified as discussing LGBTQIA+ topics, as predicted by an LLM and verified using token counts of the words lesbian, gay, bisexual, transgender, queer, intersex, and asexual. The first book is a 5th grade Social Studies textbook from our Religious collection. The relevant passage is taken from a section entitled "Gay Rights":

> "Another issue that troubled Americans was gay rights. During his time as president, Obama encouraged the growth of gay rights. He said that men should be able to marry men and that women should be able to marry women. When Obama first took office, many Americans disagreed with this view. By 2015, many agreed. Americans who disagreed faced problems. Many believed that God designed marriage to be between a man and a woman. They trusted that His plan benefited humanity. These Americans did not support gay marriage. They thought that it would harm society. For example, children in families where there is a same-sex marriage are denied having both a father and a mother. But in 2015, the Supreme Court ruled in favor of gay marriage. It had redefined marriage. Five justices voted in favor of gay marriage, but the other four justices voted against it and pointed out that this change threatened religious liberties. Some Christians who owned bakeries, flower shops, or photo studios were sued for not participating in a same-sex wedding. These businesses served all other customers, but refused to participate in same-sex weddings. Many of these business owners lost their court cases and were fined. This change in the United States raised other questions. Could religious schools continue to teach their views of marriage? Would schools and business owners come under more pressure to violate their beliefs on these issues? These questions remained unanswered."

The second book is a 3rd grade Social Studies textbook from our California collection. The relevant passage is from a section entitled "Supporting Many Different Communities":

> "Many other Californians worked hard to make changes in their communities...
>
> Harvey Milk was a political leader and a gay activist in the 1970s. Gay men are attracted to other men. Milk was ... one of the first elected officials in history to tell the public that he was gay. Milk worked hard to protect gay and lesbian rights. Lesbians are women who are attracted to other women. Milk's work also made it easier for gay, lesbian, bisexual (those attracted to both men and women), and transgender groups to stand up for their rights.
>
> Chaz Bono is a transgender activist. The word gender refers to whether someone is male or female. Transgender people feel that their gender is the opposite of what it was at birth. Bono was born female, but always felt male. He changed his gender and today lives as a man. Bono and other activists fight against discrimination and stereotypes of their community."

The passage in the religious collection highlights the tension between LGBTQIA+ rights and religious freedoms. Given that this is the only representation of LGBTQIA+ individuals in our sample, children exposed to this curriculum are only seeing people in this community posed as a threat to the freedoms of others. The passage in the collection of California

textbooks gives definitions of what it means to be a gay, lesbian, bisexual, and/or transgender individual and highlights two LGBTQIA+ activists. We did not identify any passages in the Texas state-adopted curriculum that discuss LGBTQIA+ topics, rights, or individuals.

Gender. In looking at gender presence in textbooks, we see the public school collections have higher female representation in both the text and the images of textbooks than the religious collection (Figure A.12 and Table B.3). These patterns are consistent over time; however, across all collections females are under-represented relative to their share of the population. Out of the top ten famous people and top ten first names in a collection, at most one are female, where Abigail Adams being mentioned in California and Harriet Tubman being mentioned in Texas (Tables B.2).

One could hypothesize that females are less represented in textbooks due to their historical marginalization in social studies and science fields. We examine this by disaggregating gender representation by subject. We see a distinct pattern emerge. Science and Social Studies textbooks have less female representation in the text. However, in public school Reading textbooks, females are mentioned as frequently in the text as males (Figure A.12c). This pattern may be attributed to the greater flexibility publishers have in selecting the gender of characters for Reading textbooks. Unlike history textbooks, where the narrative often revolves around historical events and figures from which women were historically marginalized (such as the consistent male dominance in the U.S. presidency) Reading textbooks are not bound by these historical constraints. Consequently, discussions in Reading materials can feature a more balanced gender representation. We do not see this same pattern in the religious collection or in images (Figures A.12c and A.12d). Indeed, females are consistently less represented in the text across all subjects within the religious school books. The closest to gender parity is found in Reading textbooks, where female-related words constitute 41% of all gendered words.

Next, we measure the contexts in which females and males are portrayed in both the text and images in two different ways: (1) through the top relative words most closely associated with females vs. males, and separately, males vs. females and (2) by testing whether females or males are more likely to be found in the context of specific domains (such as in the context of family or politics).

Table B.4 shows the top 10 relative words most closely associated with females vs. males and, separately, with males vs. females in both the text and image descriptions for each collection, along with the average VAD scores for the top words. We see that in both text and images, females are more associated with words related to appearance and males are more associated with words relating to the military and horses. Consistent with the

patterns we see for the actions related to each gender, the sentiment of the top ten relative words for females are more or equally positive, but less active and powerful on average than the top relative male words. This pattern remains present when we extend our analysis to include the top 50 relative words by gender, as shown in Figures 7a and 7b.

In Figures 7c and 7d we show the gender centeredness of several domains: *family, home, appearance, politics, sports,* and *tools*. For example, the *home* domain (which includes words such as *marriage, children, relatives, household,* etc) is more skewed towards females and the *Politics* domain (which includes words such as *government, democracy, political,* etc) is more skewed towards males. We find that in all three textbook collections, the *family* and *home* domains are skewed towards female in both the text and images. This gender skew is larger for images in the religious collection than it is for images in public school textbooks. We find a similar (though smaller) gender skew in the *politics* and *tools* domains for males in the text and images across all collections. Lastly, we see no meaningful differences by gender when examining the associations in the *sports* and *performance arts* domains in the Religious and Texas collections, but a slight skew towards females in these domains in California textbooks. We see smaller and more mixed gender skews in the *sports* and *performance arts* domains. In the text of textbooks in our religious collections, the *sports* domain is male skewed, but *performance arts* tends to either have no skew or be slightly female skewed across all collections.

VI Market and Regulatory Forces Driving Textbook Content

This section provides a conceptual framework linking textbook-market structure to curriculum content. The goal is to rationalize two empirical regularities in our data: (1) public-school textbooks adopted in California and Texas are often similar to each other on a variety of dimensions, and (2) textbooks from religious publishers are systematically more distinct along specific dimensions. Before presenting our conceptual framework, it is useful to provide some context on the setting in which these textbooks are published and adopted.

VI.A Institutional Background

Textbook adoption and purchasing in the United States is typically handled at the school-district level. However, the set of materials available to districts depends on whether they operate in an adoption state or an open (local) adoption state. Adoption states, including California and Texas for the grades and subjects in our sample, maintain a state-approved list and run formal review cycles to screen instructional materials for alignment with state standards. In these states, districts generally must select textbooks from the approved list, or else seek additional approvals to use non-listed textbooks (Doan and Kaufman, 2024).

Publishers competing in the national public-school textbook market typically start from a baseline or "national" edition and create state or district specific versions when expected sales justify the fixed costs of customization. In adoption states such as California and Texas, publishers submit materials along with documentation of standards alignment for review during the state's formal review cycle. States will then provide feedback and allow publishers to potentially revise or withdraw materials that fail compliance screens.

Once a textbook has been adopted, publishers' scope to compete on price is often constrained by procurement rules. In many adoption systems, prices are standardized through state price lists and/or multi-year state contracts that hold prices over the adoption period (often with "favored nation" or lowest-price provisions), which compresses price dispersion across approved options. For example, California maintains K–8 adopted-material price lists and imposes "lowest price" requirements and limits on post-adoption price increases, making price a comparatively second-order margin once a program clears the state's content screens. As a result, publishers tend to differentiate primarily on content rather than on pricing.

Note that the above institutional details only apply to the national public school textbook market and not the private school textbook market. State or local restrictions on educational content do not extend to private schools.

VI.B Conceptual Framework

The framework emphasizes the market and regulatory forces which influence the content of textbooks within a textbook market.

Public School Textbooks: State Standards, Costly Customization, and the Principle of Minimum Differentiation. We first consider a subset of publishers seeking broad adoption in public school districts across the nation, rather than optimizing for adoption in an individual state. A benchmark insight from spatial competition models is that when two competitors choose a position on a line and purchasers choose the closest option, each competitor has incentives to move toward the median purchaser, leading to minimum differentiation in equilibrium (Hotelling, 1929; Downs, 1957).[12] In our setting, this implies that publishers seeking scale in the public-school market will often start from a common baseline textbook designed to appeal to the median purchaser (e.g. the median school district) and then make only limited state-specific adjustments when the expected gains from customization justify the fixed cost.

[12] In the classic two-stage Hotelling model with endogenous prices, the minimum-differentiation conclusion is not generally robust. We therefore use this only as a reduced-form benchmark for environments in which price competition is muted and content positioning is the primary strategic margin. See, e.g., Eaton and Lipsey (1975); d'Aspremont et al. (1979); Economides (1984); Osborne and Pitchik (1987).

To capture how state standards affect textbook content for these broad-market public school publishers, let S_s denote the set of topics required by state s and B_s the set of prohibited or banned topics, for $s \in \{C, T\}$. A publisher chooses a baseline topic bundle A and can then add/delete topics to produce a state edition that satisfies state standards. If including a topic in the national textbook is cheaper than repeatedly customizing state editions, publishers have incentives to embed in the baseline the "safe union" of required topics:

$$A \supseteq (S_C \cup S_T) \setminus (B_C \cup B_T). \tag{1}$$

These regulatory and cost constraints yield three predictions. First, topics required in one state and not banned in the other can appear in *both* states' editions even if only one state's standards explicitly require them (positive spillovers in topic coverage). Second, topics banned by either state become unattractive for the national textbook, potentially reducing coverage even in the state that does not ban them (a "chilling" effect on baseline content). Third, when customization is required, textbook publishers will include the minimum amount of content to satisfy states while minimizing costs; we call this "Minimum Customization". Over, this framework predicts that state standards and costly customization can generate cross-state spillovers in textbook content.

Religious School Textbooks: Product Differentiation and Limited Curriculum Oversight A distinct set of publishers target private-school and homeschool purchasers who are not required to comply with state adoption standards and who may actively demand content that is differentiated from mainstream public-school materials. These publishers are therefore not competing to maximize adoption share in the same national public-school textbook market. Instead, they face a different demand distribution and a different constraint set. Standard models of product differentiation imply that when firms serve distinct segments with different preferences (or constraints), equilibrium products can diverge rather than converge (see, e.g., Fournier et al., 2020; Shaked and Sutton, 1982; Anderson et al., 1992). In our context, this segmented-market logic predicts that religious publishers will systematically differentiate their baseline topic bundles and narrative emphases along the dimensions most salient to their consumers and better align with their preferences and values, because those consumers are not constrained by state adoption rules.

Our framework is intentionally reduced-form: we do not model the decision of whether a publisher specializes in textbooks that serve religious private schools, specific states, or instead serves multiple states from a common baseline. Rather, conditional on a publisher pursuing a specific market, the framework characterizes the content incentives it faces. How-

ever, our assumptions that major public school publishers commonly participate in textbook markets across many states and textbooks published by religious publishers do not serve public schools districts are supported by anecdotal and empirical evidence. To demonstrate, Figure A.13 shows that in any given year, over half of the adopted textbooks in Texas are from publishers that also appear in the California adopted textbook lists. For California, this is even higher at over 70% in any given year.

VI.C Empirical Evidence

The predictions generated by these market and regulatory forces are consistent with our main findings in Section V. Namely (i) Texas and California textbooks are more similar to each other than we might have expected given the narrative of political polarization and (ii) textbooks from religious collection systematically differ from public school textbooks. Below, we present empirical evidence of cross-state content spillovers and show that public-school textbooks and textbooks in our religious collection are differentiated in ways consistent with differences in the beliefs and demographics of the populations they are intended to serve. These exercises are descriptive and are intended to evaluate whether the patterns in the data line up with the framework's directional predictions.

VI.C.1 Spillovers in Curriculum Standards

In this section, we examine whether discrete changes in state content standards are reflected in textbooks and whether those changes spill over across states. Over our sample period we observe multiple instances of a state altering curriculum standards for all grades in a given subject.[13] For each subject, we restrict the sample to the longest period in which only one state changes its content standards and both states subsequently adopt new textbooks. For each subject and grade, we then construct a "new-standards" dictionary from the text of the revised standards. This dictionary consists of words that appear in the updated standards for the state changing its curriculum but do not appear in the prior standards of either state. We find approximately 250 new-standards words in our Social Studies standards change (such as *abolitionist, capitalism, pollution, tuskegee, urban,* etc) 300 for Reading (such as *bilingual, folktales, idioms, interconnect, nature,* etc), and 800 for Science (such as *flowering, diet, digital, reservoirs, visual,* etc).

We measure, for each textbook-year observation, the share of words belonging to this dictionary and track how this share evolves over time by state. Changes in the average instances of these new words by state and year are shown in Figure 8. The dashed vertical line

[13]Note that changes in curriculum standards happen less frequently than changes in textbook adoptions. For example, in California, we don't observe any changes in state curriculum standards over our sample, but we do observe changes in textbook adoption.

marks the standards-change year and the solid vertical lines mark each state's subsequent textbook adoption cycle in that subject. The resulting patterns are consistent with the spillover mechanism predicted by Equation (1). In the state that revises its standards, the share of new-standards words rises sharply after the next adoption cycle, indicating a direct effect. In some cases, the state that does not revise its standards also exhibits an increase in new-standards words after its own subsequent adoption cycle, indicating a spillover effect despite no change in its standards. To more precisely estimate the effects implied by these descriptive patterns, we estimate the following textbook-year level regression:

$$y_{bcsgt} = \beta^D DirectPost_{cst} + \beta^S SpilloverPost_{cst} + \alpha_{csg} + \varepsilon_{bcsgt} \tag{2}$$

where y_{bcsgt} is the share of words from the new-standards dictionary in textbook b from collection c, subject s, grade g, in year t. $DirectPost_{cst}$ is an indicator variable which equals one in years following a change in textbook adoption for subject s in the state that also had a change in curriculum standards. $SpilloverPost_{cst}$ is an indicator variable which equals one in years following a change in textbook adoption for subject s in the state that had no change curriculum standards. The omitted category consists of observations before either state adopts new textbooks, during which time our change in standards takes place for one state. We include $collection \times subject \times grade$ fixed effects (α_{csg}) and cluster standard errors at the textbook level. Table 2 shows large and precisely estimated direct effects and smaller, but generally positive, spillovers for Reading and Social Studies textbooks. In our pooled specification, we find a direct increase of 18.5 new-standards words per 1000 words in a textbook and a spillover effect of 4.7 new-standards words for every 1000 words in a textbook. In contrast, we do not observe spillovers between Texas and California for Science textbooks. This pattern is also reflected in Figure A.3, which shows that Texas and California science textbooks have become increasingly dissimilar over time. We hypothesize this finding is partially explained by changes in science standards. In 2013, California (along with many other states) adopted the Next Generation Science Standards, whereas Texas did not[14]. This difference is visible in Figure A.8, where "Next Generation Science Standards" appears as a topic unique to the California collection and is absent from both the Texas and Religious collections. We hypothesize that this divergence in standards helps explain why we do not observe similar spillovers in Science textbooks and why Texas and California science textbooks have become less similar over time.

These exercises provide empirical evidence that changes in state standards translate

[14]Since 2013, at least 39 states and the District of Columbia have adopted the Next Generation Science Standards (NGSS) or similar standards (NGSS Lead States, 2018)

into textbook content primarily at adoption-cycle boundaries, and that revisions in one adoption state can spill over into the other states through publishers' incentives to update a common national baseline and then make only limited state-specific deviations ("minimum customization"); this spillover mechanism helps explain why California and Texas editions often move in parallel and why we observe substantial cross-state convergence in Section V.

VI.C.2 Product Differentiation by Religious Publishers

The framework in Section VI emphasizes that religious publishers do not compete in the same national public-school adoption market. Instead, they serve private-school and homeschool consumers who face different constraints and may have different preferences. If parents select schools partly to reinforce preferred beliefs, publishers may have incentives to supply curricular materials that reflect the values of the populations they serve (Wheaton, 2020; Bisin and Verdier, 2001; Tabellini, 2008). Segmented-market logic therefore predicts that religious textbooks will differ from public-school textbooks along dimensions salient to these consumers, rather than converge toward a common "median district" product. We show that several of the main differences between the Religious collection and the California and Texas collections align with differences in the demographics and beliefs of the populations these textbooks are designed to serve.

Demographics and racial representation. Our collection of religious school textbooks contain a higher share of pictured characters and references to famous individuals who are White, and depicted faces with lighter skin colors on average (Table B.2). These patterns reflect differences in student demographics across school types. Figure 9 shows that students in Religious private schools and homeschool settings are substantially more likely to be White than students in California and Texas public schools (e.g., roughly 70% White in Religious private schools and among homeschoolers versus about 21% in California public schools and 26% in Texas public schools). This demographic gap is large enough that even a publisher with no intrinsic preference over representation may rationally tailor examples and imagery toward the modal student audience in its target market.

Beliefs and Values If religious publishers are differentiating to match their consumers' values, we should see differences not only in who is represented, but also in which norms and beliefs are emphasized. To probe this mechanism, we use the restricted geocoded General Social Survey (GSS) to compare beliefs of (i) practicing religious respondents nationwide,[15] (ii) non-practicing respondents in California, and (iii) non-practicing respondents in Texas.

[15]We define "practicing" as attending religious services approximately once per week. Many respondents who do not meet this threshold still self-identify as religious; the intent is to isolate a group for whom religion is plausibly a central organizing identity.

Figure 11a uses a long-running GSS module (1983–2018) asking respondents to choose the most important trait to teach children. Practicing religious respondents place substantially more weight on teaching children to obey and less weight on teaching children to think for themselves than non-practicing respondents in California and Texas; differences in "helping others" and "working hard" are comparatively smaller. Textbooks exhibit a qualitatively similar pattern: Religious-collection textbooks are more likely to contain obedience-related values and much less likely to contain critical-thinking values than California and Texas textbooks. While obedience and critical-thinking language appear on a small fraction of pages in all collections (under 6%), the direction of the differences aligns with the survey evidence, consistent with purposeful positioning toward a different demand segment. At the same time, Religious-collection textbooks place substantially more emphasis on explicit "caring" and "hard work" values than public-school collections, suggesting that differentiation is not limited to a small set of contentious topics but extends to broader and more explicit moral and character instruction as shown in Figure A.5. This is also consistent with survey data indicating that parents who choose to enroll their children in a private school commonly report a preference for more character values education (Catt and Rhinesmith, 2016).

A similar pattern appears for other salient beliefs. Figure 10 shows large differences in views on evolution and prayer in schools: practicing religious respondents are far more likely to reject evolution and to oppose bans on prayer in schools than non-practicing respondents in California and Texas. These belief gaps mirror the content differences we document in the religious collection, where religious topics are much more prevalent (Figure 3) and prayer is among the most prominent topics in reading materials (Figure A.4). Figure 10 also shows that practicing religious respondents are more likely to endorse traditional gender norms (e.g., agreement that men are better suited for politics and that women belong in the home). Consistent with this, Religious-collection textbooks exhibit lower female representation (Table B.3) and systematically portray females in contexts that are less active and less dominant than males in both text and images relative to public school textbooks (Figure 7).

Taken together, these descriptive exercises show that religious publishers appear to differentiate along dimensions that are predictably valued by their consumers and that track measurable differences in demographics and beliefs, rather than simply reflecting idiosyncratic publisher choices. More broadly, the same market forces that generate cross-state convergence in the public-school adoption market can coexist with substantial divergence once publishers target different, unregulated segments with distinct preferences.

VII Conclusion

Curricula play a vital role in transmitting knowledge, shaping values, and constructing collective memory. Growth in U.S. school choice programs has increased public funds directed to religious schools, but little is known about what is taught and how it compares to public school curriculum. Our study analyzes textbooks from public schools, religious private schools, and home schools using computational methods and advances from artificial intelligence to assess how people, topics, and values are represented over time. Contrary to widespread narratives of political polarization, we find few substantial differences between public school textbooks from Texas and California. We show that this is consistent with textbook market forces when publishers are constrained by state standards and customization is costly. In contrast, religious school textbooks show less representation of females, more depictions of light-skinned individuals, and differing treatments of topics such as evolution and religion, including in science content. We show this is consistent with differences in demographics and beliefs between families who send their kids to public schools versus religious private schools or homeschool. Notable similarities also appear across all three collections: both public and religious textbooks rarely address LGBTQIA+ topics and individuals, have minimal representation of Asian, Indigenous, and Latine individuals, depict females more positively but in less active or powerful roles than males.

It is important to note these methods perform better when assessing overall portrayals rather than examining the nuances of individual representations or measuring the salience of particular messages, suggesting a potential gap that could be addressed through manual content analysis as evidenced by the differences in portrayal uncovered in our examination of the passages discussing topics related to evolution and LGBTQIA+ individuals.

In our setting, we focus our analysis on printed curriculum, which could differ from the values that children are otherwise exposed to in classrooms. Similarities in textbooks across different educational settings do not necessarily guarantee that children receive similar educational experiences in the classroom. Pedagogy plays a crucial role in shaping how material is taught, and differences in teacher training, beliefs, and classroom practices can lead to substantially different lessons even when the same textbook is used. However, while standardizing classroom instruction requires a highly prescriptive and often impractical implementation strategy, textbook adoption remains a more feasible lever for policy intervention. Importantly, prior research has shown that changes in curriculum standards alone can influence student outcomes, suggesting that printed curriculum matters and the importance of analyzing textbook content remains despite potential differences in pedagogy.

Our paper demonstrates how advances in computational social science and artificial intelligence tools can be used to measure differences in educational materials and other sources of text and image data. Prior textbook content analysis studies have focused on cross-sectional analyses of gender and racial representation in text alone. We build on and extend prior research on textbook content analysis by analyzing both the text and images of a novel panel of elementary school textbooks used in public and religious settings from 1980 to the present. Our longitudinal approach allows us to measure changes in the presence and portrayal of people, topics, and values over time and across educational settings.

In the midst of an increase in educational gag orders and statewide changes in the content standards of public school curriculum, our work shines a light on what messages children are receiving from informal and non-traditional curricula outside the public school system which could have long-term implications for children's beliefs and decision-making.

References

Abdulkadiroğlu, Atila, Parag A Pathak, and Christopher R Walters. 2018. "Free to choose: Can school choice reduce student achievement?" American Economic Journal: Applied Economics, 10(1): 175–206.

Adukia, Anjali, Alex Eble, Emileigh Harrison, Hakizumwami Birali Runesha, and Teodora Szasz. 2023. "What we teach about race and gender: Representation in images and text of children's books." The Quarterly Journal of Economics, 138(4): 2225–2285.

Adukia, Anjali, Callista Christ, Anjali Das, and Ayush Raj. 2022a. "Portrayals of race and gender: Sentiment in 100 years of children's literature." Proceedings of the 5th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies, 20–28.

Adukia, Anjali, Patricia Chiril, Callista Christ, Anjali Das, Alex Eble, Emileigh Harrison, and Hakizumwami Birali Runesha. 2022b. "Tales and tropes: Gender roles from word embeddings in a century of children's books." Proceedings of the 29th International Conference on Computational Linguistics, 3086–3097.

Adukia, Anjali, Richard Hornbeck, Daniel Keniston, and Benjamin Lualdi. 2025. "The social construction of race during Reconstruction."

Allen, Ann M, Daniel N Allen, and Gary Sigler. 1993. "Changes in sex-role stereotyping in Caldecott Medal award picture books 1938—1988." Journal of Research in Childhood Education, 7(2): 67–73.

Anderson, Simon P., André de Palma, and Jacques-François Thisse. 1992. Discrete Choice Theory of Product Differentiation. MIT Press.

Arold, Benjamin W. 2024. "Evolution vs. creationism in the classroom: The lasting effects of science education." The Quarterly Journal of Economics.

Arold, Benjamin W., Ludger Woessmann, and Larissa Zierow. 2025. "Can Schools Change Religious Attitudes?" Journal of Human Resources. Article ID: 0723-13023R2.

Ash, Elliott, Ruben Durante, Maria Grebenschikova, and Carlo Schwarz. 2021. "Visual representation and stereotypes in news media."

Bartholet, Elizabeth. 2020. "Homeschooling: Parent rights absolutism vs. child rights to education & protection." Ariz. L. Rev., 62: 1.

Bazzi, Samuel, Andreas Ferrara, Martin Fiszbein, Thomas P Pearson, and Patrick A Testa. 2023. "The Confederate diaspora."

Bickford, John Holden, and Lori A Knoechel. 2017. "The historical representation of Native Americans within primary-and intermediate-level trade books." The Councilor: A Journal of the Social Studies, 78(2): 4.

Bisin, Alberto, and Thierry Verdier. 2001. "The economics of cultural transmission and the dynamics of preferences." Journal of Economic theory, 97(2): 298–319.

Braden, Eliza G, and Sanjuana C Rodriguez. 2016. "Beyond mirrors and windows: A critical content analysis of Latinx children's books." Journal of Language and Literacy Education, 12(2): 56–83.

Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. 2017. "Semantics derived automatically from language corpora contain human-like biases." Science, 356(6334): 183–186. Publisher: American Association for the Advancement of Science.

Cantoni, Davide, Yuyu Chen, David Y Yang, Noam Yuchtman, and Y Jane Zhang. 2017. "Curriculum and ideology." Journal of Political Economy, 125(2): 338–392.

Catt, Andrew D, and Evan Rhinesmith. 2016. "Why parents choose: A survey of private school and school choice parents in Indiana." Friedman Foundation for Educational Choice.

Charlesworth, Tessa ES, Aylin Caliskan, and Mahzarin R Banaji. 2022. "Historical representations of social groups across 200 years of word embeddings from Google Books." Proceedings of the National Academy of Sciences, 119(28): e2121798119.

Clark, Roger. 2007. "From margin to margin? Females and minorities in Newbery and Caldecott Medal-winning and Honor books for children." International Journal of Sociology of the Family, 263–283.

Crawfurd, Lee, Christelle Saintis-Miller, and Rory Todd. 2024. "Sexist textbooks: Automated analysis of gender bias in 1,255 books from 34 countries." PloS one, 19(10): e0310366.

Crisp, Thomas, Roberta Gardner, and Matheus Almeida. 2018. "The all-heterosexual world of children's nonfiction: A critical content analysis of LGBTQ identities in Orbis Pictus Award books, 1990–2017." Children's Literature in Education, 49(3): 246–263.

d'Aspremont, Claude, Jean Jaskold Gabszewicz, and Jacques-François Thisse. 1979. "On Hotelling's "Stability in Competition"." Econometrica, 47(5): 1145–1150.

Deckman, Sherry L, Ellie Fitts Fulmer, Keely Kirby, Katharine Hoover, and Abena Subira Mackall. 2018. "Numbers are just not enough: A critical analysis of race, gender, and sexuality in elementary and middle school health textbooks." Educational Studies, 54(3): 285–302.

Deming, David J, Justine S Hastings, Thomas J Kane, and Douglas O Staiger. 2014. "School choice, school quality, and postsecondary attainment." American Economic Review, 104(3): 991–1013.

Doan, Sy, and Julia Kaufman. 2024. "What Role Do States Play in Selecting K-12 Textbooks?" The Journal of the National Association of State Boards of Education. Accessed 2025-12-30.

Downs, Anthony. 1957. An Economic Theory of Democracy. Harper.

Eaton, B. Curtis, and Richard G. Lipsey. 1975. "The Principle of Minimum Differentiation Reconsidered: Some New Developments in the Theory of Spatial Competition." The Review of Economic Studies, 42(1): 27–49.

Economides, Nicholas. 1984. "The Principle of Minimum Differentiation Revisited." European Economic Review, 24(3): 345–368.

EdChoice. 2023. "Fast facts on school choice." Accessed on March 9, 2024.

Fournier, Gaëtan, Karine Van der Straeten, and Jörgen Weibull. 2020. "Spatial competition with unit-demand functions." arXiv preprint arXiv:2001.11422.

Friedman, Jonathan, and James Tager. 2022. "Educational gag orders: Legislative restrictions on the freedom to read, learn, and teach."

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou. 2018. "Word embeddings quantify 100 years of gender and ethnic stereotypes." Proceedings of the National Academy of Sciences, 115(16): E3635–E3644.

Hotelling, Harold. 1929. "Stability in Competition." The Economic Journal, 39(153): 41–57.

Jamison, Peter, Laura Meckler, Prayag Gordy, C Ence Morse, and C Alcantara. 2023. "Home schooling's rise from fringe to fastest-growing form of education." Washington Post.

Kaleem, Jaweed. 2016. "Homeschooling Without God." theatlantic.com. Accessed 2025-12-30.

Kamel, Donia, and L Guillermo Woo-Mora. 2023. "Skin tone penalties: Bottom-up discrimination in football."

Katz, Phyllis A., and Jennifer A. Kofkin. 1997. "Race, gender, and young children." In Developmental Psychopathology: Perspectives on Adjustment, Risk, and Disorder. , ed. Suniya S. Luthar, Jacob A. Burack, Dante Cicchetti and John R. Weisz, 51–74. Cambridge University Press.

Keith, Verna M, and Carla R Monroe. 2016. "Histories of colorism and implications for education." Theory Into Practice, 55(1): 4–10.

Kessen, William, and Sally Weiskopf. 1976. "Color vision and hue categorization in young human infants." Journal of Experimental Psychology: Human Perception and Performance, 2(1): 115–129.

Kinzler, Katherine D, and Elizabeth S Spelke. 2011. "Do infants show social preferences for people differing in race?" Cognition, 119(1): 1–9.

Klein, Rebecca. 2017. "Voucher schools championed by Betsy DeVos can teach whatever they want. turns out they teach lies."

Koss, Melanie D. 2015. "Diversity in contemporary picturebooks: A content analysis." Journal of Children's Literature, 41(1): 32.

Koss, Melanie D, and Kathleen A Paciga. 2020. "Diversity in Newbery Medal-winning titles: A content analysis." Journal of Language and Literacy Education, 16(2): n2.

Koss, Melanie D, Nancy J Johnson, and Miriam Martinez. 2018. "Mapping the diversity in Caldecott books from 1938 to 2017: The changing topography." Journal of Children's Literature, 44(1): 4–20.

Kozlowski, Austin C, Matt Taddy, and James A Evans. 2019. "The geometry of culture: Analyzing the meanings of class through word embeddings." American Sociological Review, 84(5).

Kunzman, Robert. 2010. "Homeschooling and religious fundamentalism." International Electronic Journal of Elementary Education, 3(1): 17–28.

Kunzman, Robert, and Milton Gaither. 2020. "Homeschooling: An updated comprehensive survey of the research." Other Education-the journal of educational alternatives, 9(1): 253–336.

Lucy, Li, Dorottya Demszky, Patricia Bromley, and Dan Jurafsky. 2020. "Content analysis of textbooks via natural language processing: Findings on gender, race, and ethnicity in Texas US history textbooks." AERA Open, 6(3).

Ludwig, Jens, and Sendhil Mullainathan. 2024. "Machine learning as a tool for hypothesis generation." The Quarterly Journal of Economics, 139(2): 751–827.

Manzini, Thomas, Yao Chong Lim, Yulia Tsvetkov, and Alan W Black. 2019. "Black is to criminal as Caucasian is to police: Detecting and removing multiclass bias in word embeddings."

Michalopoulos, Stelios, and Christopher Rauh. 2024. "Movies." National Bureau of Economic Research Working Paper 32220.

Michalopoulos, Stelios, and Melanie Meng Xue. 2021. "Folklore." The Quarterly Journal of Economics, 136(4): 1993–2046.

Mohammad, Saif M. 2018. "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words." Melbourne, Australia.

Musaddiq, Tareena, Kevin Stange, Andrew Bacher-Hicks, and Joshua Goodman. 2022. "The pandemic's effect on demand for public schools, homeschooling, and private schools." Journal of Public Economics, 212.

National Center for Education Statistics. 2019. "School Choice in the United States: 2019." U.S. Department of Education, Institute of Education Sciences NCES 2019-106. Private school enrollment data from fall 2015.

National Center for Education Statistics. 2022. "Homeschooled children and reasons for homeschooling." Condition of Education, Retrieved December 18th, 2023.

NGSS Lead States. 2018. "Science Assessment Challenges." Achieve, Inc. Next Generation Science Standards Brief.

Osborne, Martin J., and Carolyn Pitchik. 1987. "Equilibrium in Hotelling's Model of Spatial Competition." Econometrica, 55(4): 911–922.

Riley, Emma. 2022. "Role models in movies: The impact of Queen of Katwe on students' educational attainment." The Review of Economics and Statistics, 1–48.

Rodriguez, Pedro L, Arthur Spirling, and Brandon M Stewart. 2023. "Embedding regression: Models for context-specific description and inference." American Political Science Review, 117(4): 1255–1274.

Sailor, Angela, Lindsey M Burke, Anne Segal, and Adam Kissel. 2021. "Civics studies: Why they matter, what parents and teachers think, and how they can reclaim truth." Heritage Foundation.

Sallabaş, Muhammed Eyyüp. 2013. "Analysis of narrative texts in secondary school textbooks in terms of values education." Educational Research and Reviews, 8(8): 361.

Scaramanga, Jonny, and Michael J Reiss. 2018. "Accelerated Christian Education: A case study of the use of race in voucher-funded private Christian schools." Journal of Curriculum Studies, 50(3): 333–351.

Shaked, Avner, and John Sutton. 1982. "Relaxing Price Competition Through Product Differentiation." The Review of Economic Studies, 49(1): 3–13.

Stevenson, Betsey, and Hanna Zlotnik. 2018. "Representations of men and women in introductory economics textbooks." Vol. 108, 180–85.

Szasz, Teodora, Emileigh Harrison, Ping-Jung Liu, Ping-Chang Lin, Hakizumwami Birali Runesha, and Anjali Adukia. 2022. "Measuring representation of race, gender, and age in children's books: Face detection and feature classification in illustrated images." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision.

Tabellini, Guido. 2008. "Institutions and culture." Journal of the European Economic Association, 6(2-3): 255–294.

Voth, Hans-Joachim, and David Yanagizawa-Drott. 2024. "Image(s)." CEPR Discussion Papers, , (19219).

Wade, Rahima C. 1993. "Content analysis of social studies textbooks: A review of ten years of research." Theory & Research in Social Education, 21(3): 232–256.

Wheaton, Brian. 2020. "Laws, beliefs, and backlash." Working Paper.

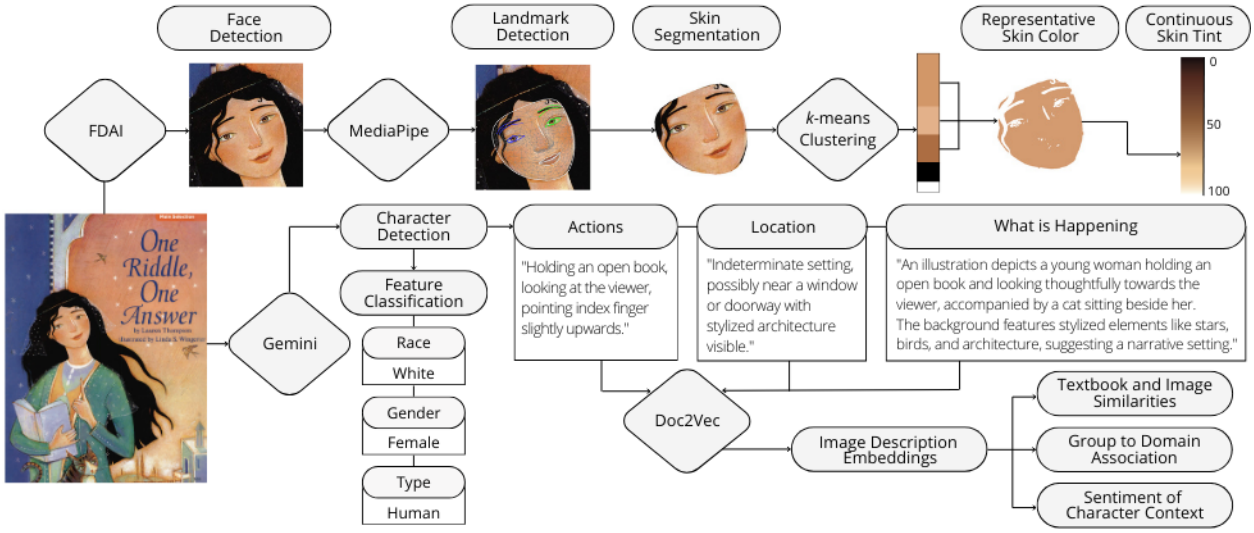
Williamson, Danielle Graves, and Jennifer Withrow. 2025. "Southern Academies: The Proliferation of All-White Private Schools After Brown and Their Legacy for Students." Working paper.

Yu, Amy Zhao, Shahar Ronen, Kevin Hu, Tiffany Lu, and César A Hidalgo. 2016. "Pantheon 1.0, a manually verified dataset of globally famous biographies." Scientific Data, 3(1): 1–16.

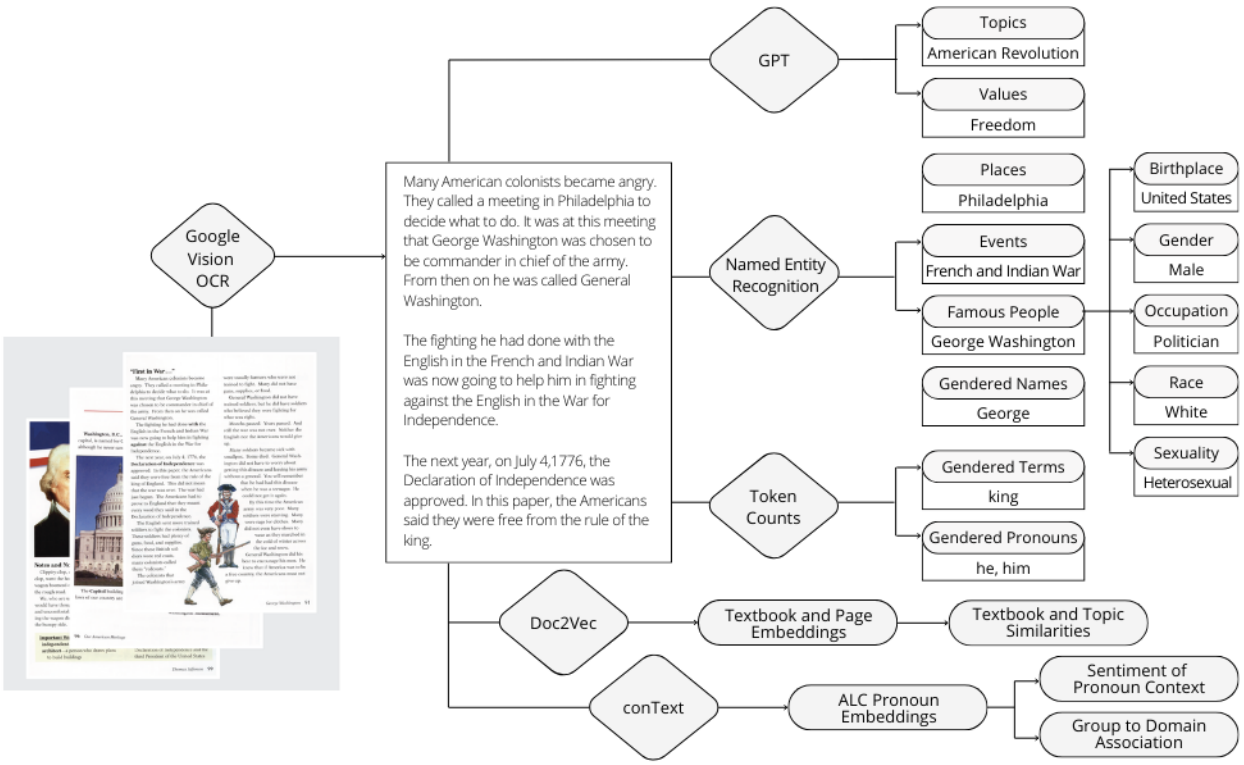
VIII Main Figures

Figure 1: Data Extraction Pipelines: Image and Text Analysis

(a) Image Analysis Pipeline

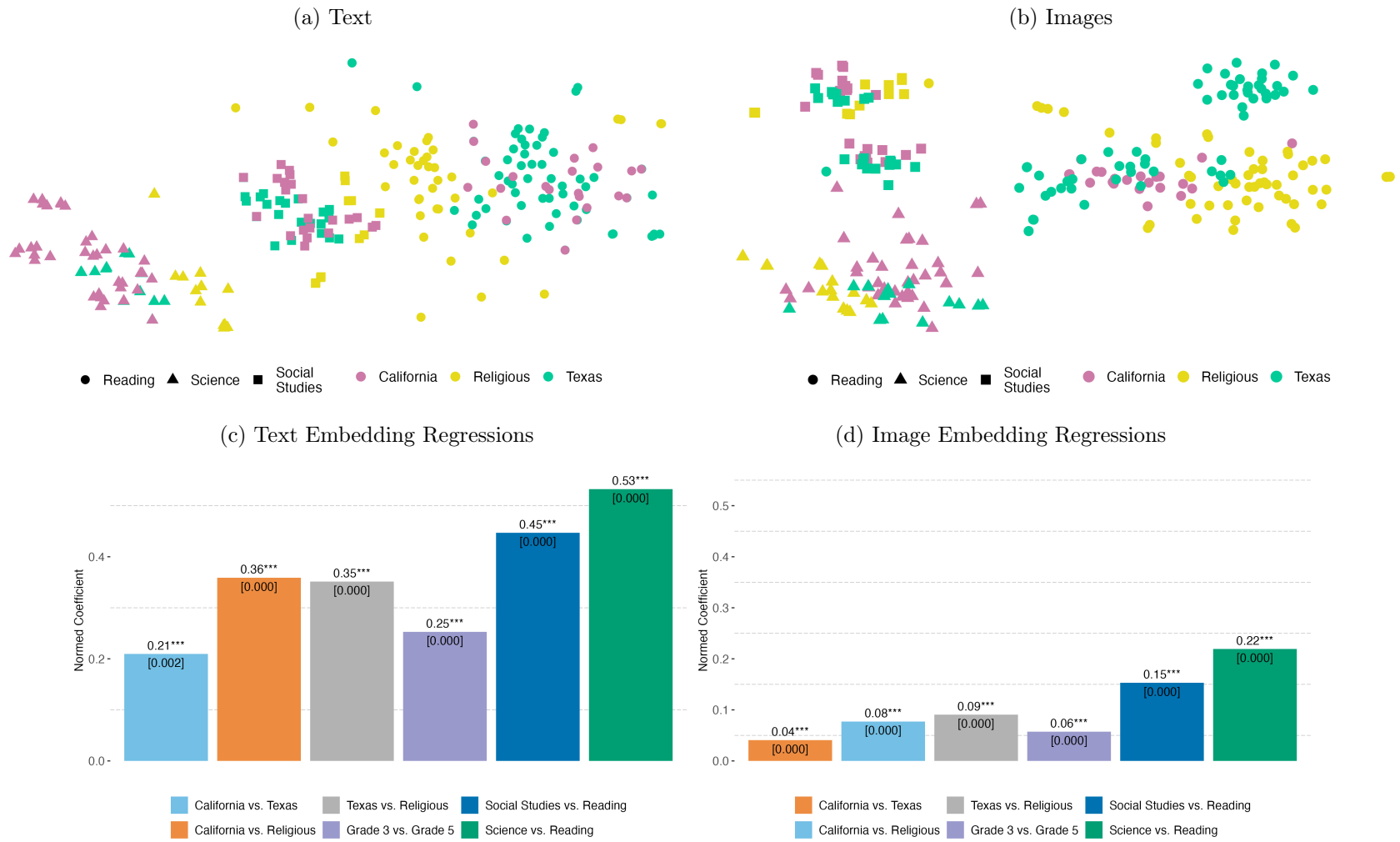


(b) Text Analysis Pipeline



Notes: This figure illustrates the steps in our image analysis (top) and text analysis (bottom) pipelines beginning with the raw scan of each page in our sample of books.

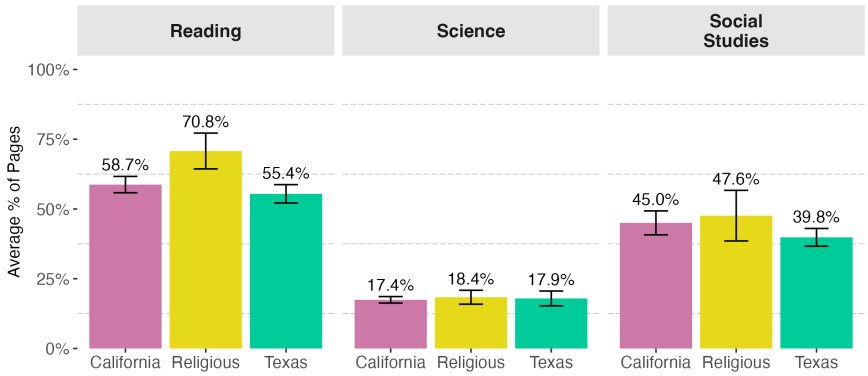
Figure 2: Textbook Embeddings



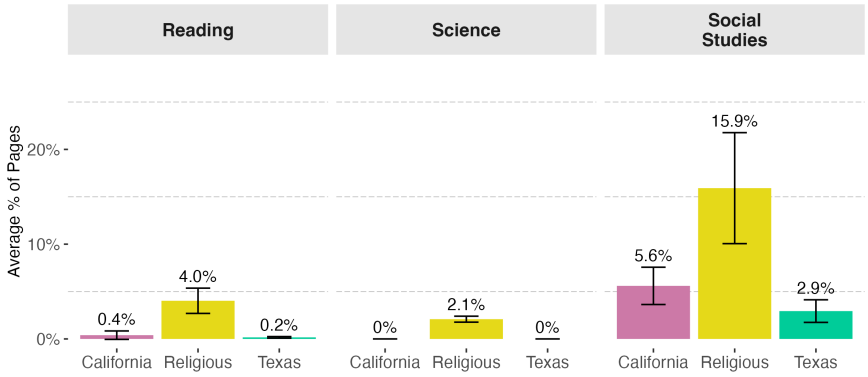
Notes: In panel (a) and (c), we show analyses using textbook-level embeddings estimated using the text from each textbook. In panels (b) and (d), we show textbook-level embeddings constructed by averaging across image embeddings in a textbook. Each image embedding is estimated from an AI generated description of a detected character's setting, location, and actions as described in Section IV. In panels (a) and (b), each observation represents a book from a specific collection (color) and subject (shape). Their content, originally represented in a 300-dimensional space, is reduced to 2D via t-SNE while retaining meaningful information. Book representations should be interpreted in terms of their relative position to other observations. In (c) and (d), we report normed coefficients from embedding regressions in the spirit of Rodriguez et al. (2023) as described in Appendix Section D. These normed coefficients measure the overall magnitude of semantic difference in embedding space between groups, aggregated across all embedding dimensions. Larger values indicate greater semantic differences between groups, while values near zero indicate little systematic separation in embedding space. We estimate two separate regressions; one where the Texas collection is the excluded group and one where the Religious collection is the excluded group. Statistical significance is assessed using permutation tests.

Figure 3: Values

(a) Average Percent of Pages Containing Character Values by Subject



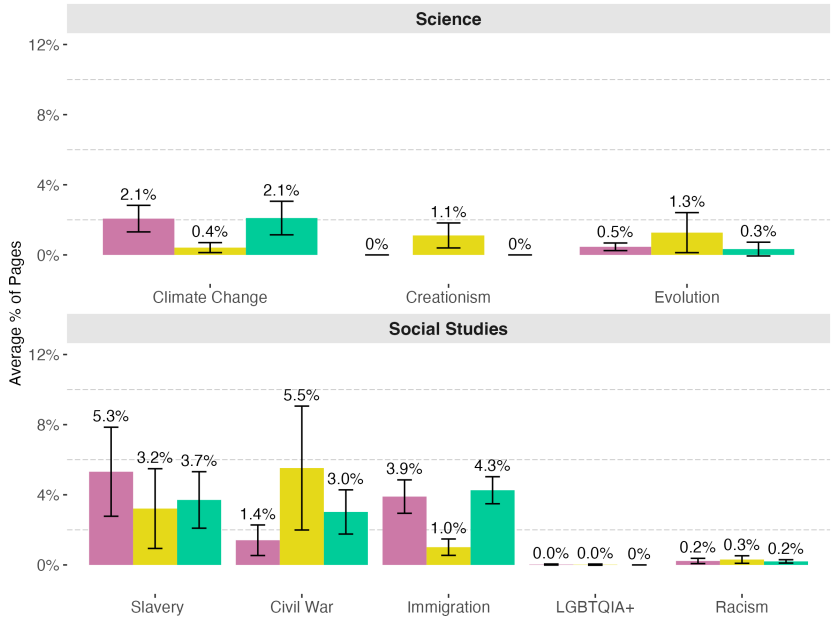
(b) Average Percent of Pages Containing Religious Values by Subject



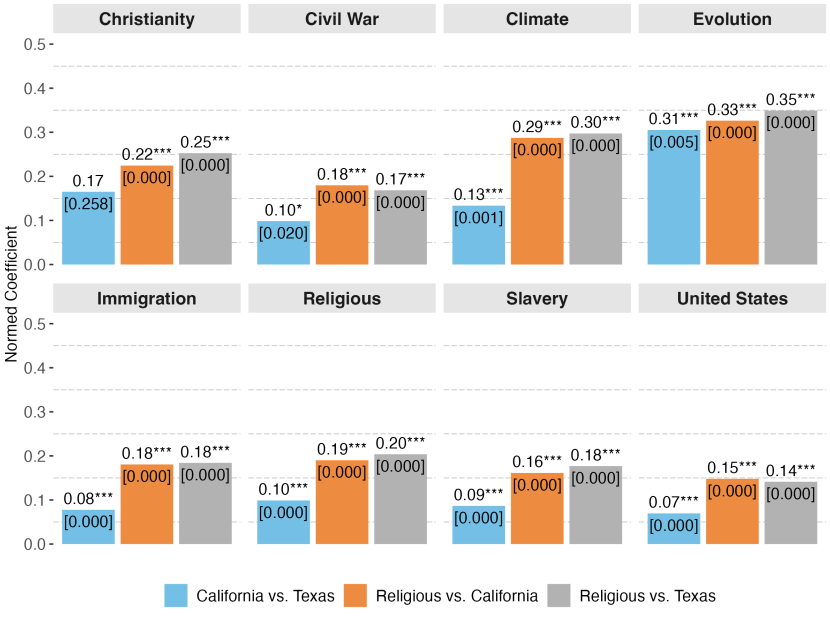
Notes: Panel (a) reports the average share of pages that contain character values by collection and subject and Panel (b) shows the average share of pages that contain religious values by collection and subject. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and subject so that each grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure 4: High-Profile Topics

(a) Average Percent of Pages Containing High-Profile Topics



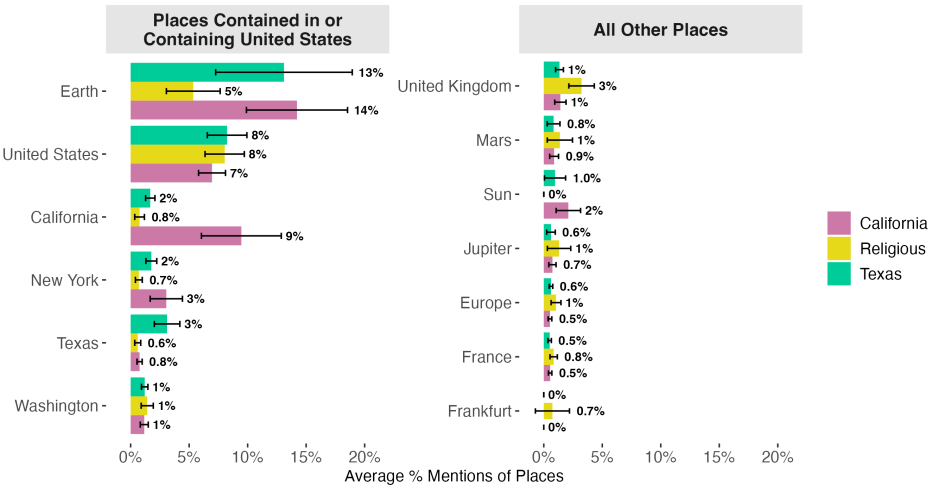
(b) Topic Embedding Regressions



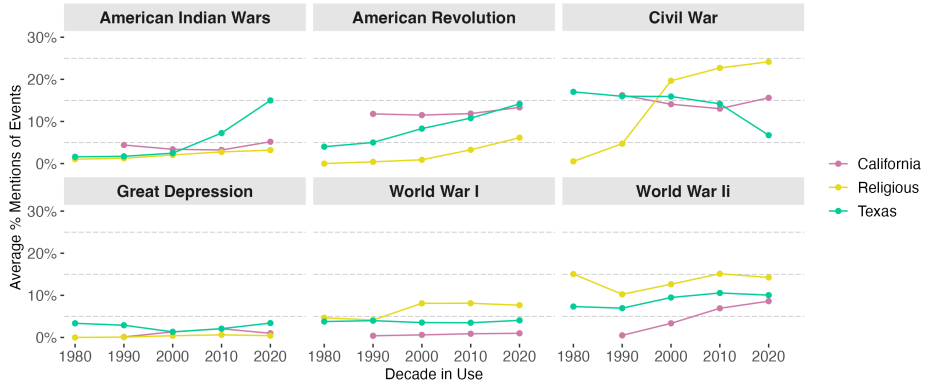
Notes: Panel (a) reports average share of pages containing a specified topic in a given collection and subject. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and subject so that each grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals. Panel (b) reports normed coefficients from a page-level embedding regression in the spirit of Rodriguez et al. (2023) as described in Appendix Section D except the regression is at the page-embedding level instead of the textbook-embedding level and for each regression we subset to pages discussing a given topic to test semantic differences in topics between collections. These normed coefficients measure the overall magnitude of semantic difference in embedding space between groups, aggregated across all embedding dimensions. Larger values indicate greater semantic differences between groups, while values near zero indicate little systematic separation in embedding space. We estimate two separate regressions; one where the Texas collection is the excluded group and one where the Religious collection is the excluded group. Statistical significance is assessed using permutation tests.

Figure 5: Geography, History, and Birthplaces

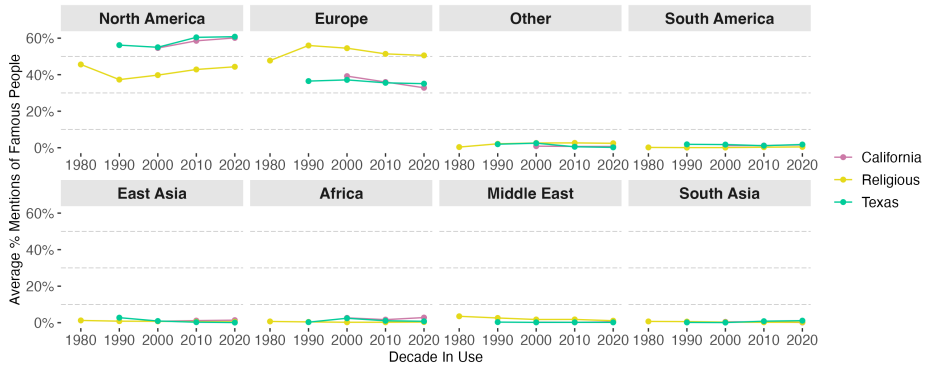
(a) Geography



(b) Historical Events



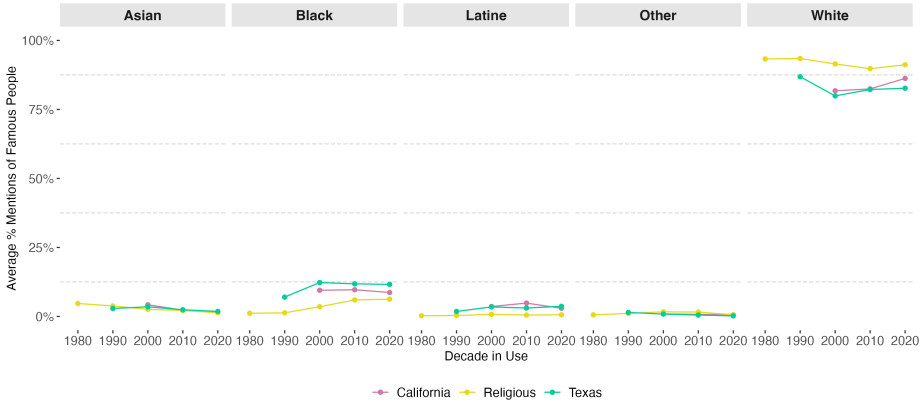
(c) Birthplace of Famous Individuals



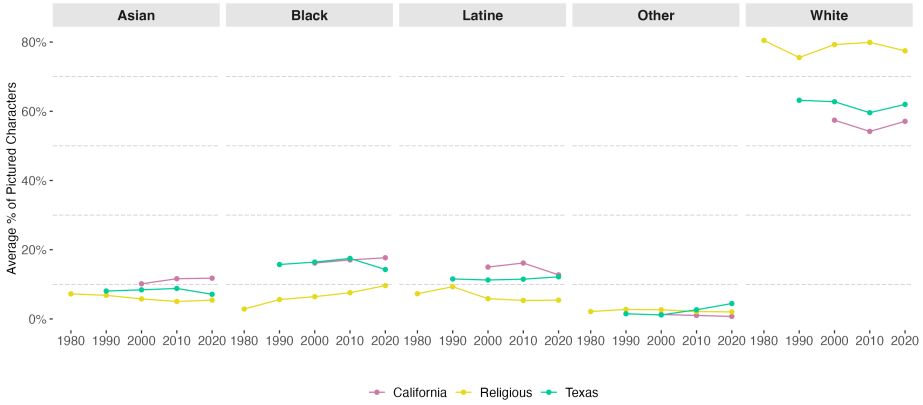
Notes: Panel (a) shows the average percentage of places mentioned accounted for by the most frequently mentioned places by collection. Panel (b) shows the average percentage of events mentioned accounted for by the most frequently mentioned events by collection over time. Panel (c) shows the average percentage of famous people mentioned by their country of birth for each collection over time. We determine mentions of places, events, and people using Named Entity Recognition. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection or collection and decade so that each subject-grade–year cell receives equal weight. Standard errors clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure 6: Representation of Race and Skin Color in Textbooks

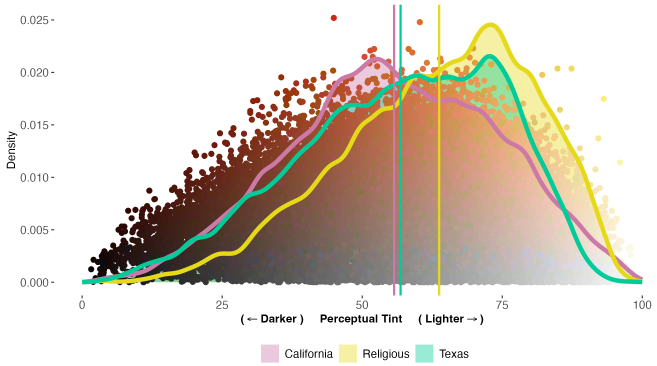
(a) Race (Text)



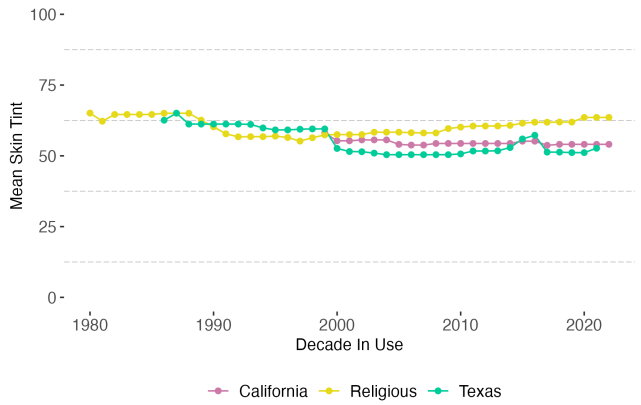
(b) Race (Images)



(c) Skin Tint Distributions



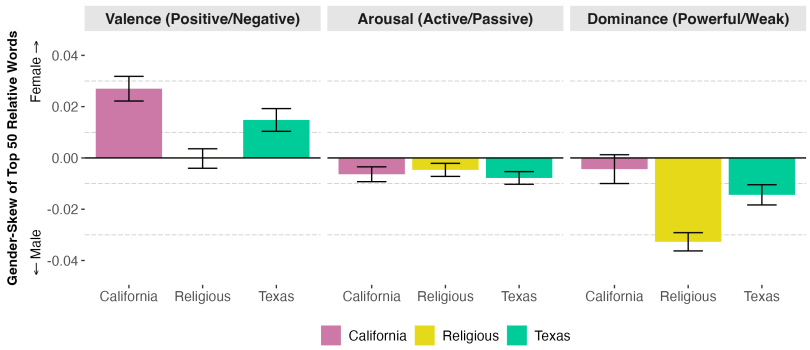
(d) Skin Tint Over Time



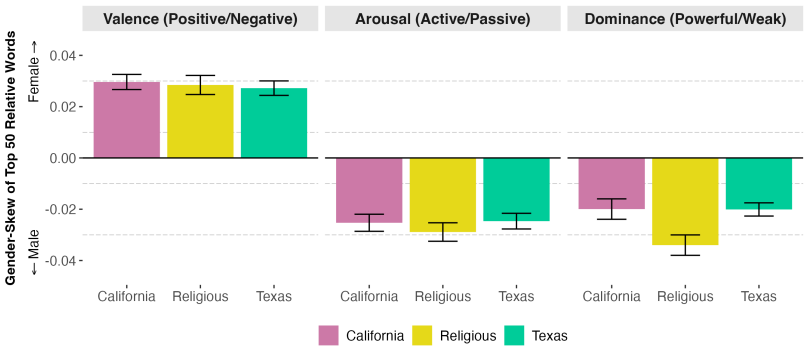
Notes: This figure shows the presence of predicted putative race and skin color. Panel (a) shows the average percent of mentions of famous people in the text that are of a specific race in each collection by decade in use. Panel (b) shows the average percent of pictured characters classified as a specific race in each collection by decade in use. Panels (c) and (d) show our analysis of the representative skin colors of detected faces in images that have human skin colors (polychromatic skin colors where $R \geq G \geq B$) or are from monochromatic images. Panel (c) shows the distribution of skin color tint for faces detected. The x-axis for both the density and scatter plot is skin tint, but the y-axis for the scatterplot is the vibrancy of each skin color. The mean for each distribution is denoted with a dashed line. Panel (d) shows the average skin tint over time. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and decade so that each subject-grade–year cell receives equal weight.

Figure 7: Gender Skew of Domains and Sentiment (VAD Scores)

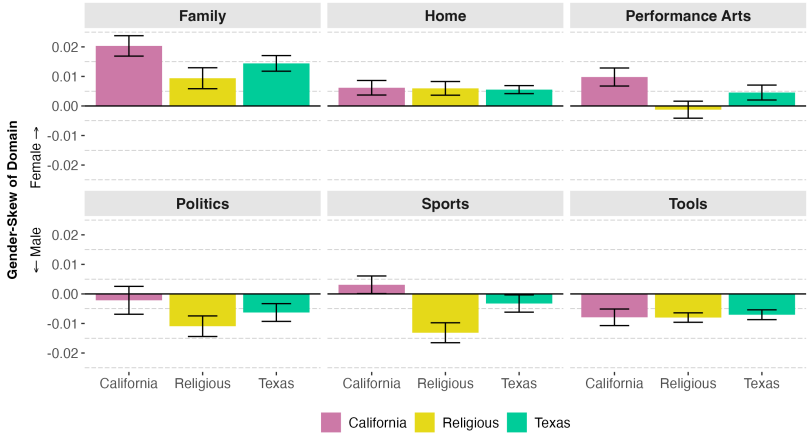
(a) Sentiment via VAD (Text)



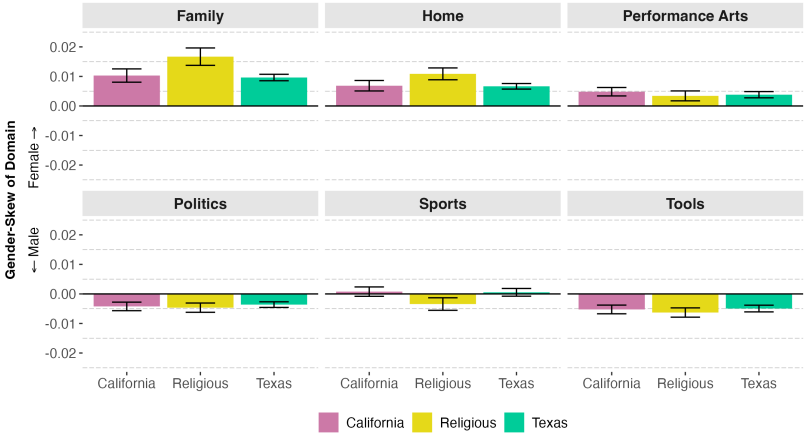
(b) Sentiment via VAD (Images)



(c) Gender Skew by Domain (Text)

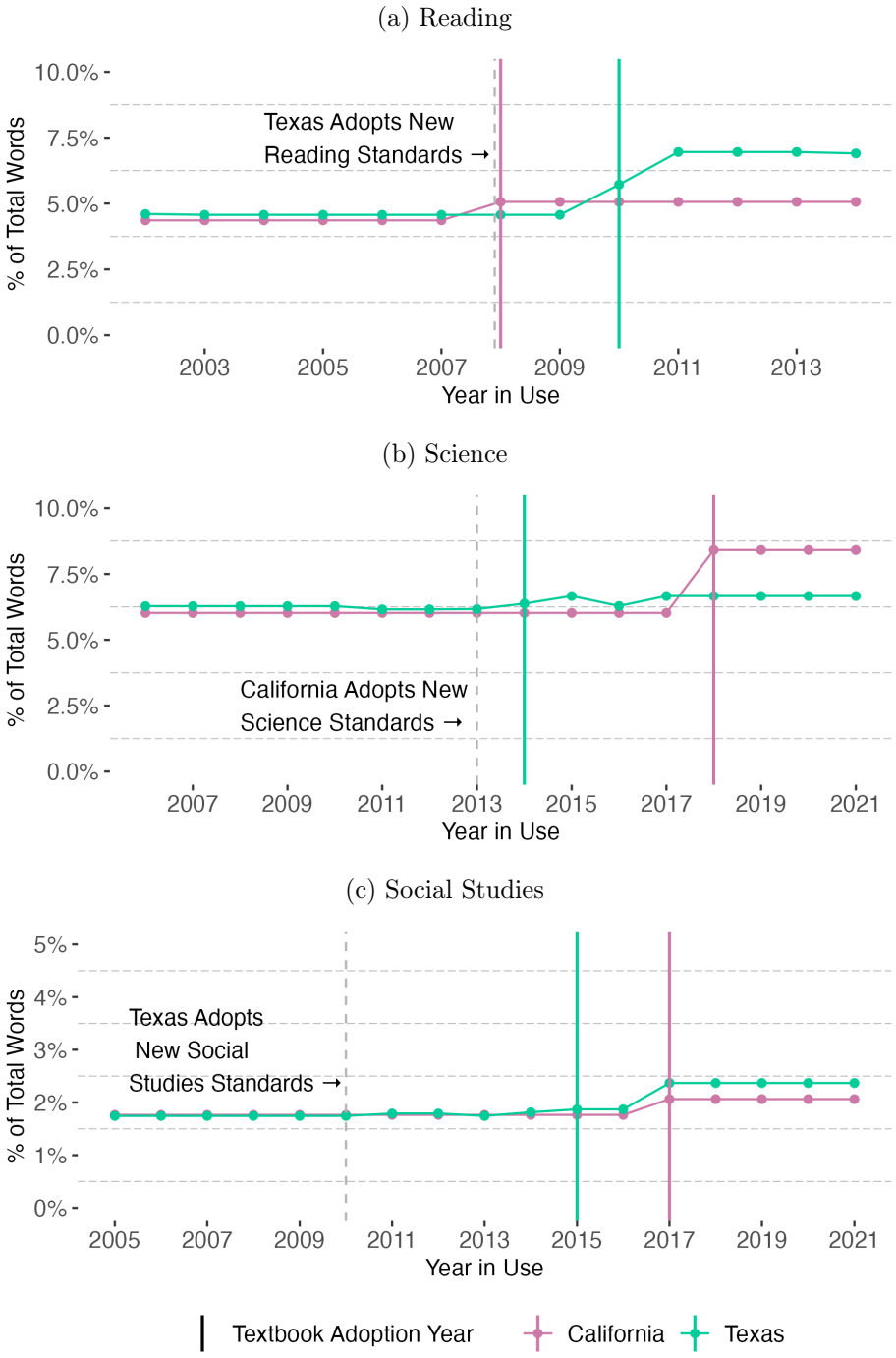


(d) Gender Skew by Domain (Images)



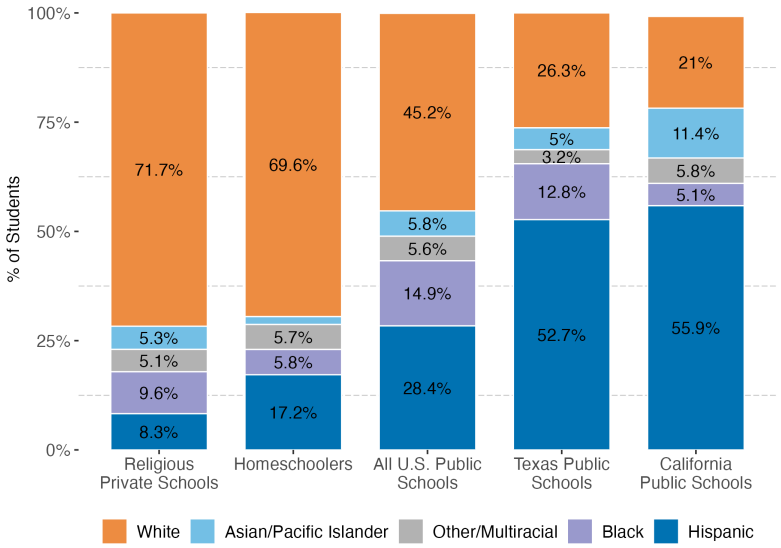
Notes: Bars report collection-specific coefficients estimating differences between males and females from embedding-based regressions; error bars denote 95% confidence intervals, with standard errors clustered at the textbook level. Panels (a) and (b) show gender differences in the average Valence, Arousal, and Dominance (VAD) scores of the 50 top relative words drawn from the NRC VAD lexicon. In panel (a), these relative words are identified from pronoun-instance text embeddings; in panel (b), they are identified from embeddings of model-generated image descriptions. Positive values indicate that female contexts are associated with words that are more positive, active, or powerful than male contexts within the same collection. Panels (c) and (d) show differences between males and females in average cosine similarity to pre-specified domain word sets (family, home, performance arts, politics, sports, and tools). Positive values indicate that a domain is more strongly associated with female than male contexts; negative values indicate the reverse. We call this "gender skew". More details available in Appendix D. All regressions control for subject, grade, and year in use.

Figure 8: Standards-Driven Vocabulary in Textbooks



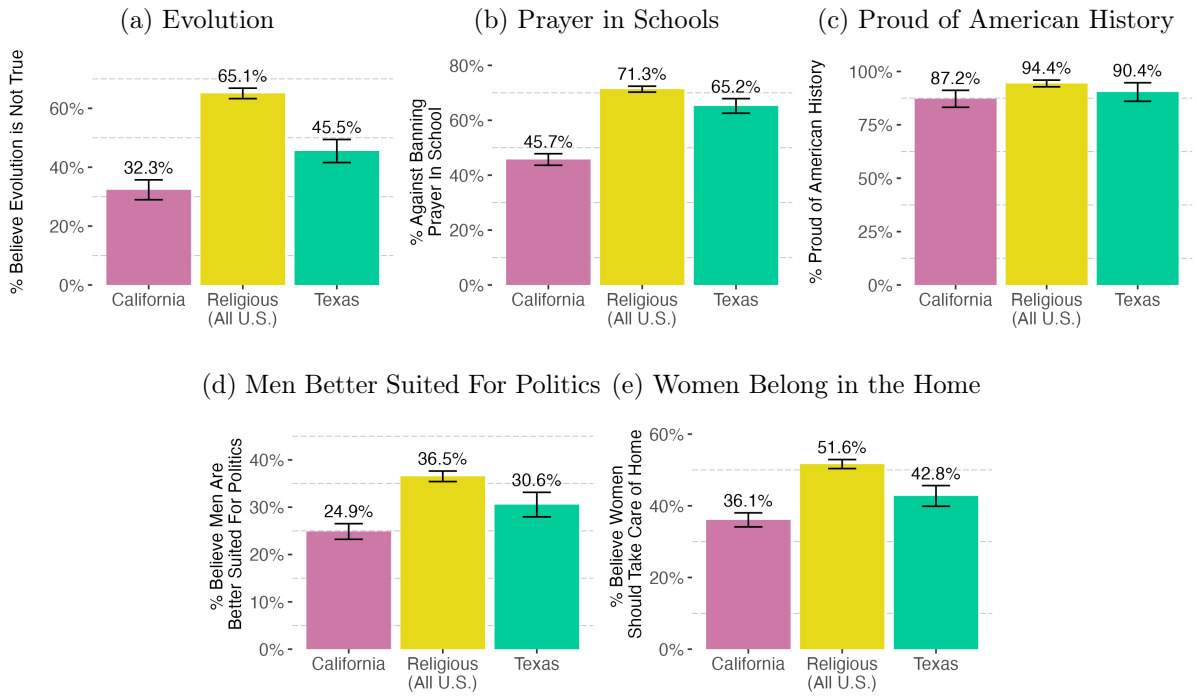
Notes: Each panel plots, by year in use, the share of words in California and Texas textbooks that belong to a dictionary of "new-standards" terms. For each subject and grade, we define this dictionary as new words that appear in the adopting state's revised standards but not in its prior standards and not in the other state's pre-period standards; we then count these terms in textbook text and average the resulting shares across grades by collection. The dashed vertical line marks the standards-revision year, and the solid vertical lines mark each state's subsequent textbook adoption cycle with different states denoted in different colors. The fact that these terms often rise in the non-adopting state after its adoption cycle is consistent with cross-state spillovers. Note that new terms are new to the state standards, not necessarily new to textbooks. For each subject, we subset to the longest time period for which we have one state standard change, and one textbook adoption cycle for each state after the standards change.

Figure 9: Racial/Ethnic Composition of K–12 Students by School Type



Notes: Data compiled from several sources. Data on religious private school students from NCES PSS 2021–22, Digest Table 205.30 (Other Religious, Non-Catholic); Data on public school students from NCES CCD 2021–22, Digest Table 203.70; Data on homeschoolers from NCES PFI-NHES 2019, Digest Table 206.10 (homeschoolers). Homeschooler Asian/Pacific Islander category is back-calculated from reported totals and includes American Indian and Alaskan Native students.

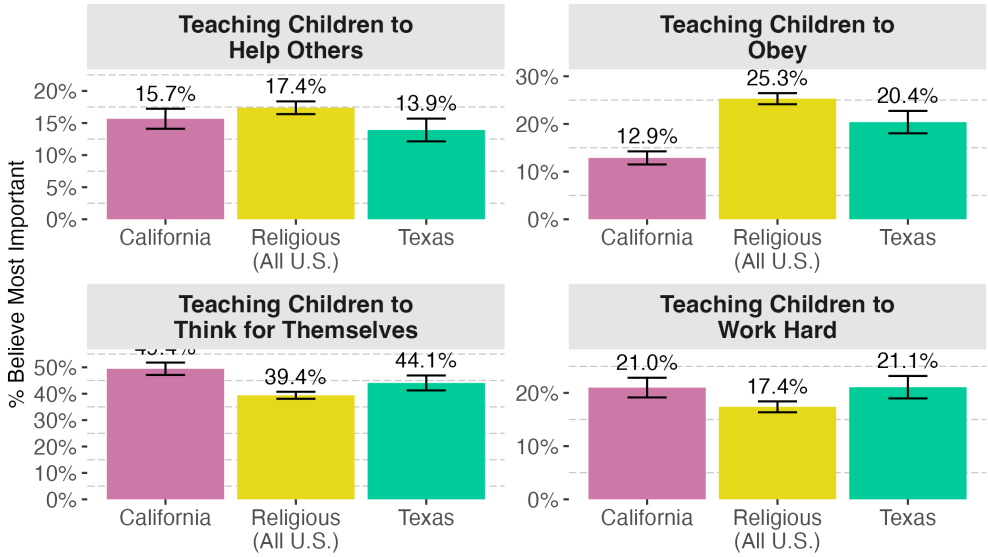
Figure 10: Beliefs by State and Religiosity



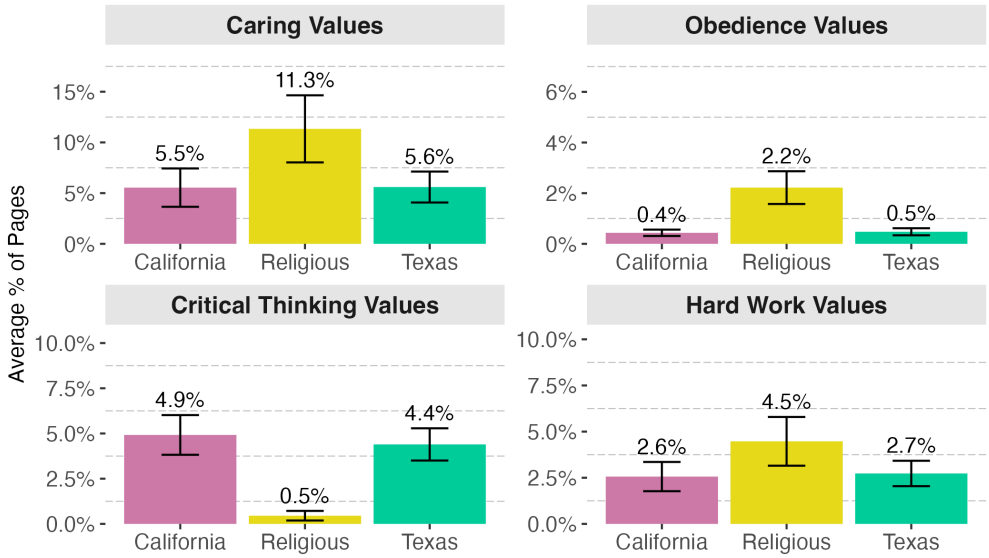
Notes: Bars report survey-weighted means from pooled General Social Survey cross-sections, using all years in which each question was asked; error bars denote 95% confidence intervals that account for the GSS survey design. The "Religious (All U.S.)" group consists of respondents nationwide who attend religious services weekly or more often. "California (Non-Religious)" and "Texas (Non-Religious)" consist of respondents in those states who do not attend religious services weekly; thus, weekly attenders in California and Texas are included in the religious group rather than the state groups. In panel (a), "Believe Evolution Is Not True" is a composite measure constructed from available GSS evolution/creationism items.

Figure 11: Character Values Categories

(a) GSS Survey on Most Important Things to Teach Children



(b) Average Percent of Pages Containing Values Related to GSS Survey



Notes: Panel (a) reports survey-weighted means from pooled General Social Survey (GSS) cross-sections, using all years in which this question was asked; error bars denote 95% confidence intervals that account for the GSS survey design. The "Religious (All U.S.)" group consists of respondents nationwide who attend religious services weekly or more often. "California (Non-Religious)" and "Texas (Non-Religious)" consist of respondents in those states who do not attend religious services weekly; thus, weekly attenders in California and Texas are included in the religious group rather than the state groups; however, respondents who identify as religious but do not attend services weekly are included in the group of people from Texas and California. Panel (b) shows the average share of pages that contain specific values related the the GSS question in Panel (a). Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection so that each subject-grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

IX Main Tables

Table 1: Summary Statistics

	California (1)	Religious (2)	Texas (3)
Collection Totals			
Number of Reading textbooks	22	55	60
Number of Science textbooks	36	17	14
Number of Social Science textbooks	22	13	22
Number of grade 3 textbooks	42	50	55
Number of grade 5 textbooks	38	35	41
Total number of textbooks	80	85	96
Total number of pages	29,262	19,814	39,235
Total number of faces	23,664	14,065	22,040
Total number of pictured characters	43,596	27,342	56,221
Sample Period			
Earliest year in use	1999	1980	1986
Latest year in use	2022	2022	2021
Average number of years textbooks were in use	7	13	8
Weighted Average: Page Content			
Average number of topics per page	8	8	8
Average number of values per page	6	6	6
Weighted Average: Skin Color			
Average skin tint of faces	54	59	53
Weighted Average Percent: Putative Race of Pictured Characters			
Average percent of pictured characters classified as Asian	10%	6%	7%
Average percent of pictured characters classified as Black	15%	6%	15%
Average percent of pictured characters classified as Latine	14%	6%	11%
Average percent of pictured characters classified as White	51%	72%	57%
Average percent of pictured characters classified as other race	1%	2%	2%
Average percent of pictured characters classified as unsure race	9%	9%	8%
Weighted Average Percent: Putative Race of Famous People			
Average percent of famous people classified as Asian	3%	3%	3%
Average percent of famous people classified as Black	9%	3%	10%
Average percent of famous people classified as Indigenous	1%	1%	1%
Average percent of famous people classified as Latine	4%	0%	3%
Average percent of famous people classified as White	83%	92%	83%
Weighted Average Percent: Gender of Pictured Characters			
Average percent of pictured characters classified as female	25%	17%	26%
Average percent of pictured characters classified as male	43%	45%	44%
Average percent of pictured characters classified as unsure gender	32%	38%	30%
Weighted Average Percent: Gendered Words and Famous People			
Average percent of female gendered words	39%	26%	42%
Average percent of famous people classified as female	15%	9%	18%
Weighted Average Percent: Character Type			
Average percent of pictured characters classified as human	72%	63%	72%

Notes: Columns report the three U.S. textbook collections in our sample. The "Collection Totals" panel reports counts aggregated across unique textbooks in each collection For all averages, we first compute the mean for textbooks within each collection × subject × grade × year-in-use cell and then average these cell means within each collection, so that each subject–grade–year cell receives equal weight. This procedure ensures that textbooks used over longer periods are given greater weight and that equal emphasis is placed on each subject and grade within in a given year. Average number of topics and values per page on conditional on having at least one topic/value identified on a page.

Table 2: Direct Effects and Cross-State Spillovers from State Standards Revisions

	Dependent Variable: % of Total Textbook Words			
	All Subjects	Reading	Science	Social Studies
	(1)	(2)	(3)	(4)
Direct Effect	0.0185***	0.0188***	0.0220***	0.0042**
	(0.0023)	(0.0031)	(0.0035)	(0.0017)
Spillover	0.0047***	0.0075***	0.0028	0.0027**
	(0.0009)	(0.0015)	(0.0019)	(0.0011)
Fixed-effects				
Collection-Grade-Subject	Yes			
Collection-Grade		Yes	Yes	Yes
Observations	761	285	292	184
Within R^2	0.476	0.435	0.605	0.356

Notes: This table reports estimates of Equation (2). The unit of observation is a textbook-year. For each subject, we restrict the sample to the longest period in which only one state changes its curriculum standards and both states subsequently adopt new textbooks. We construct a subject and grade specific "new-standards" dictionary consisting of words that appear in the revised standards of the state changing its standards but not in the prior standards of either state. The dependent variable is the share of textbook words from this dictionary. "Direct Effect" indicates the coefficient on post-adoption textbook-years in the standards-changing state, and "Spillover" indicates the coefficient on post-adoption textbook-years in the other state. The omitted category consists of textbook-year observations before either state adopts new textbooks. A positive and significant "Direct Effect" indicates that revised standards translate into textbook content at adoption-cycle boundaries, while a positive and significant "Spillover Effect" indicates that these changes diffuse across states through publishers' cross-market production decisions. In the pooled specification, the direct effect implies an increase of about 18.5 new-standards words per 1,000 textbook words, while the spillover effect implies an increase of about 4.7 new-standards words per 1,000 textbook words. Standard errors are clustered at the textbook level. *p<0.1; **p<0.05; ***p<0.01.

Online Appendix

Separation of Church and State Curricula? Examining Public and Religious Private School Textbooks*

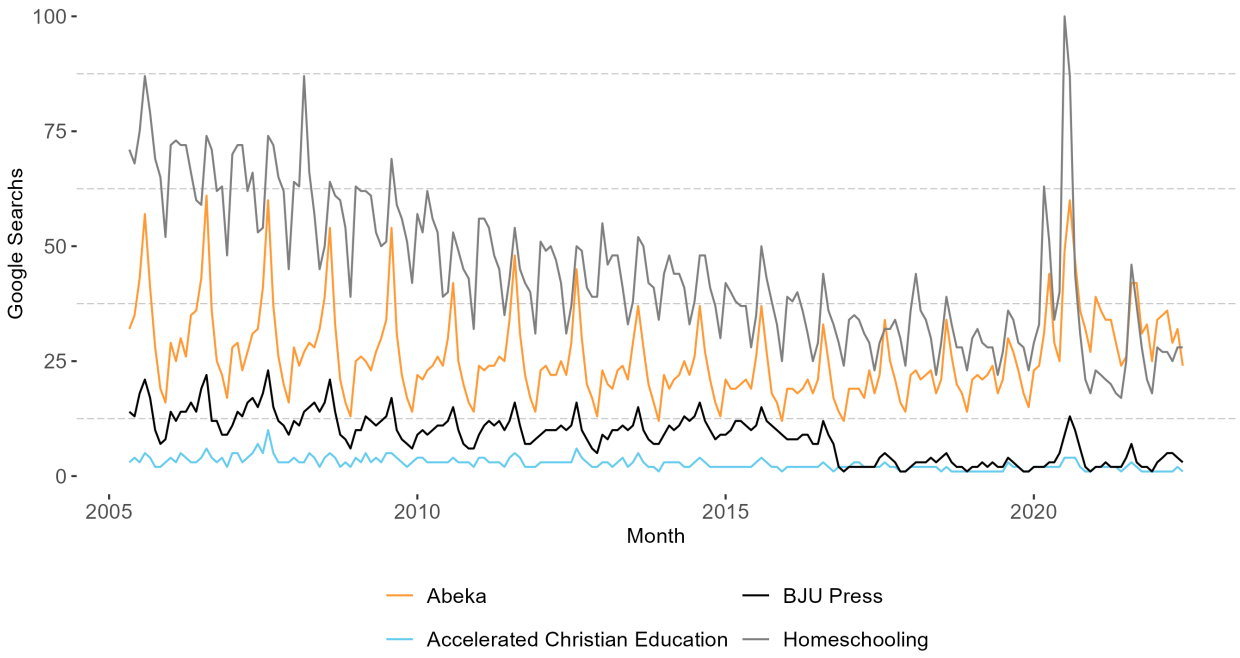
March 2026

Anjali Adukia, University of Chicago, MiiE Lab, IZA, NBER
Emileigh Harrison, University of Chicago, MiiE Lab, Inclusive Economy Lab

*Contact: adukia@uchicago.edu, emileigh.harrison@austin.utexas.edu

A Appendix Figures

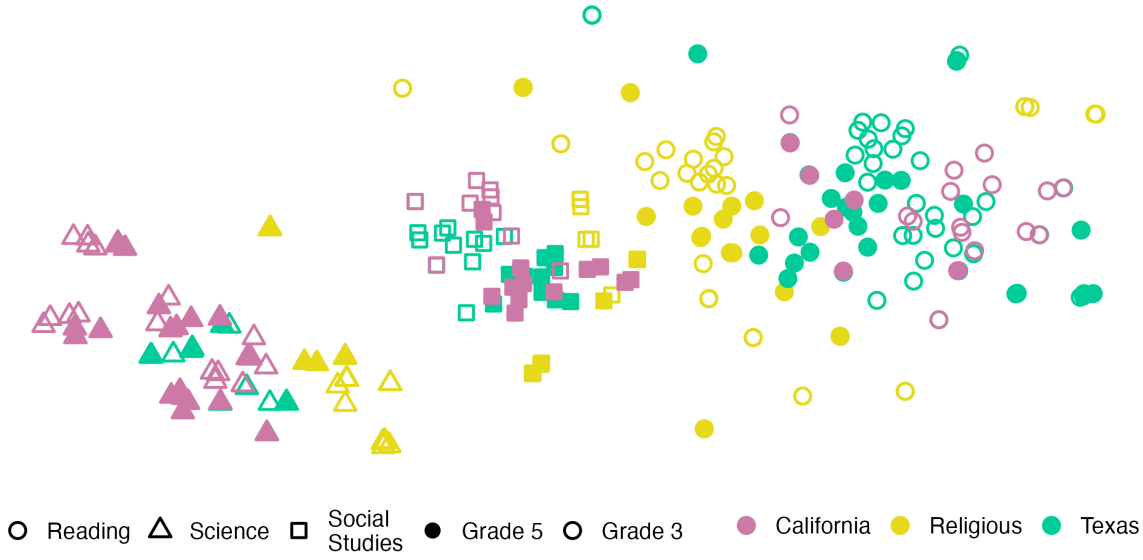
Figure A.1: Google Search Interest Over Time — Homeschooling and Religious Publishers



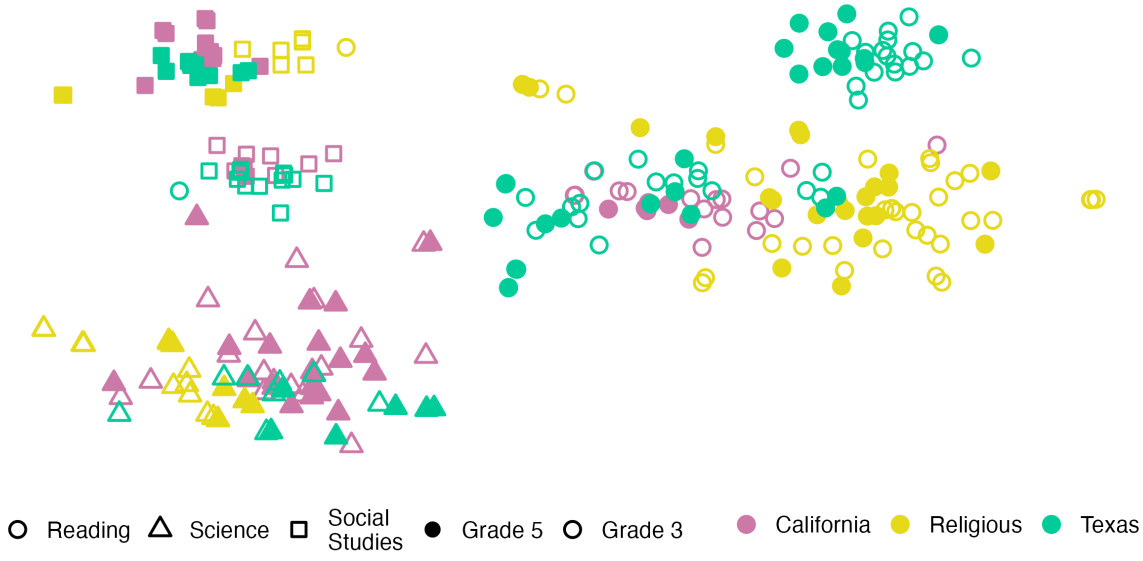
Notes: This figure shows Google search interest from Google Trends data for the search term homeschool and each of the education publishing companies in the religious school textbook collections; Abeka, Acceleration Christian Education (ACE), and BJU Press (BJU). Note that we use the company topics created by Google Trends instead of the search terms when looking at these three publishers. This figure shows the increase in search interest for homeschooling and Abeka curriculum after the start of the COVID-19 crisis.

Figure A.2: Textbook Embeddings

(a) Text



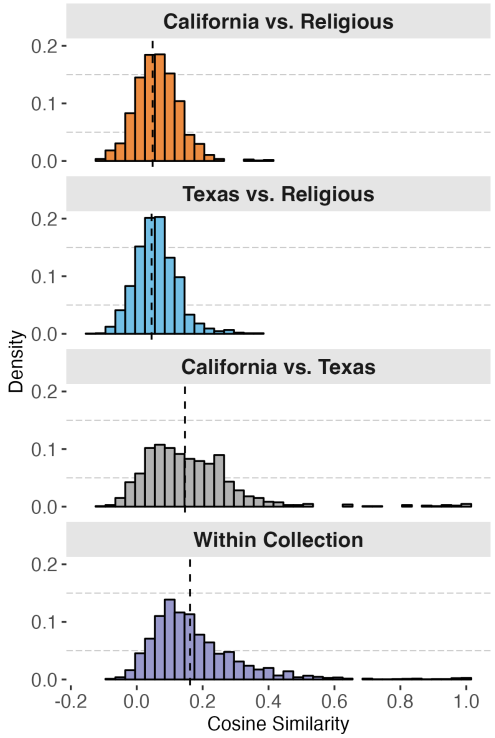
(b) Images



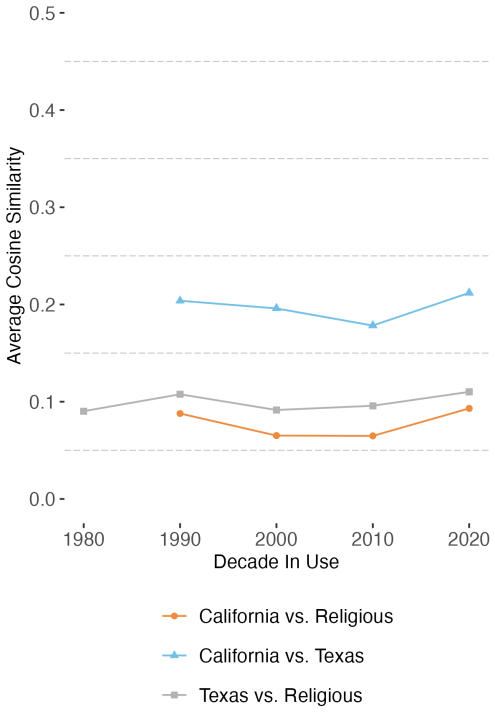
Notes: Each observation represents a book from a specific collection (color), subject (shape), and grade (outline). The filled shapes indicate 5th grade textbooks and hollow shapes indicate 3rd grade textbooks. Their content, originally represented in a 300-dimensional space, is here reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and book representations should be interpreted in terms of their relative position to other observations. Among collections, pink signifies California, yellow signifies Religious, and green signifies Texas. Among subjects, the squares (○) signify Reading, the triangles (△) signify Science, and the circles (□) signify Social Studies.

Figure A.3: Average Cosine Similarity Between Book Embeddings by Collection Pairs

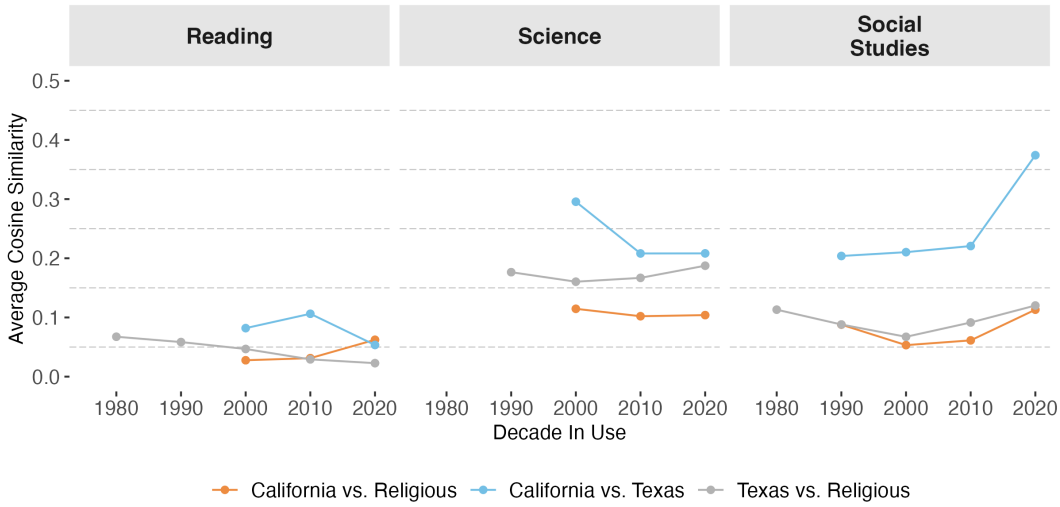
(a) Histogram of Similarities Between Book Embeddings by Collection Pairs



(b) Average Similarity Between Book Embeddings by Collection Pairs over Time



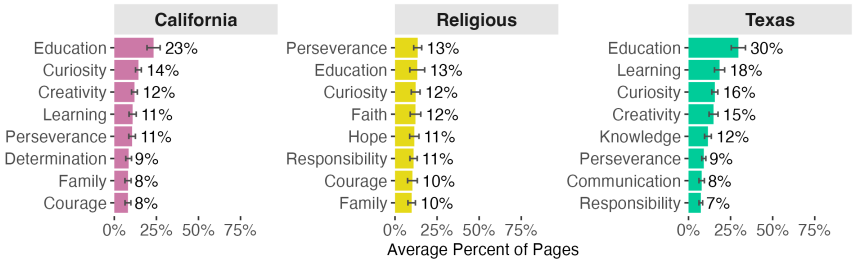
(c) Average Similarity Between Book Embeddings by Collection Pairs and Subject over Time



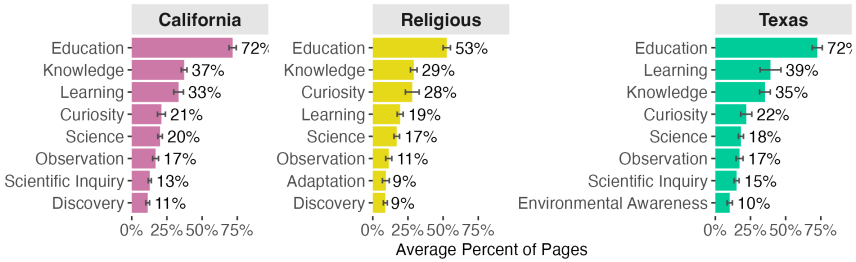
Notes: In Figure (a), we plot a histogram of the cosine similarities between two books that share the same subject, grade, and year in use. We separate the densities by collection pairs (e.g. one California textbook and one Texas textbook) and plot within collection book pair similarities for comparison. Note that each book pair is repeated for every year those two books are both in use. In Figure (b), we average the cosine similarities of book pairs for each decade and collection pair; this average is weighted equally across subject, grade, and year in use. Figure (c) follows the same approach as Figure (b) except that we calculate the averages separately by subject.

Figure A.4: Top Values and Topics by Collection

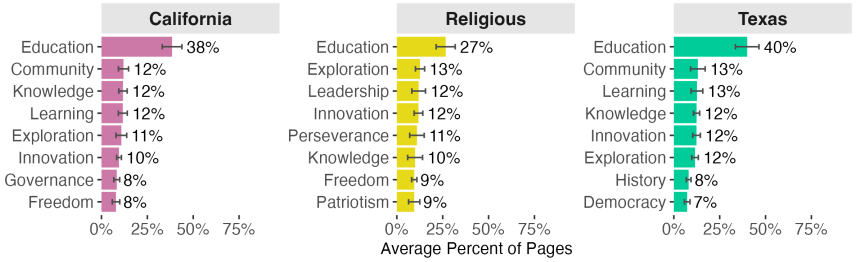
(a) Top Values in Reading



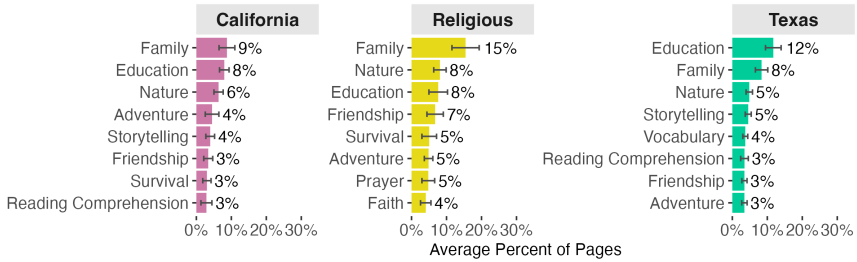
(b) Top Values in Science



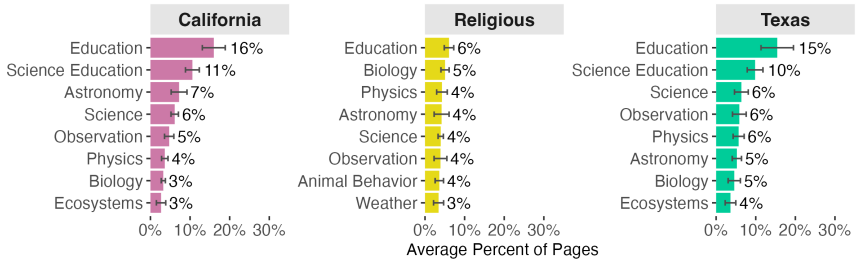
(c) Top Values in Social Studies



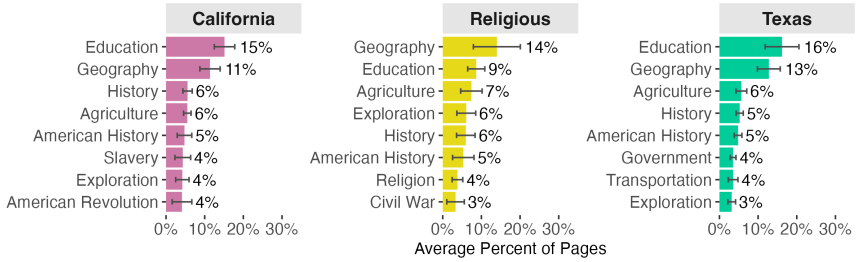
(d) Top Topics in Reading



(e) Top Topics in Science



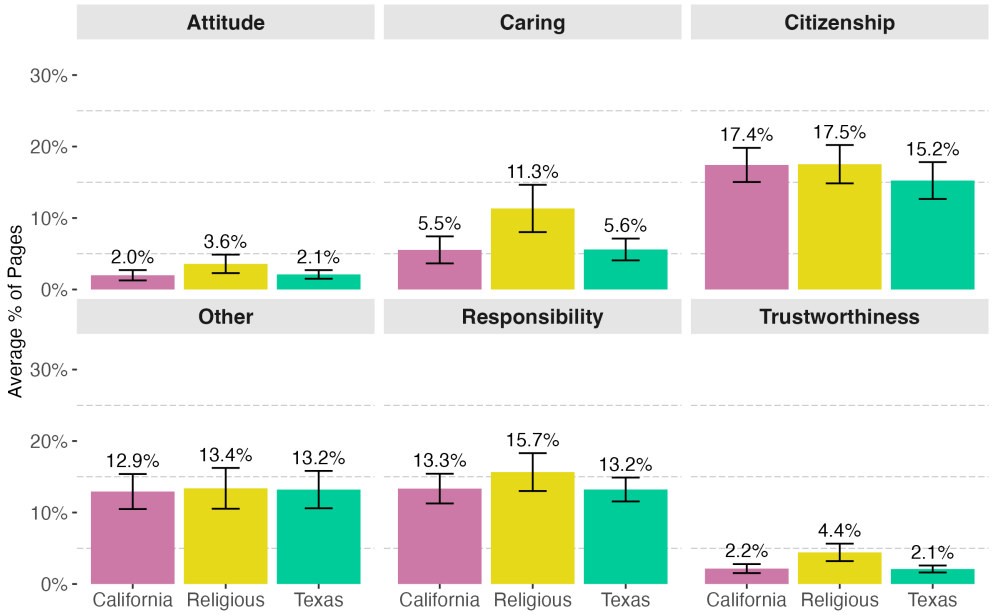
(f) Top Topics in Social Studies



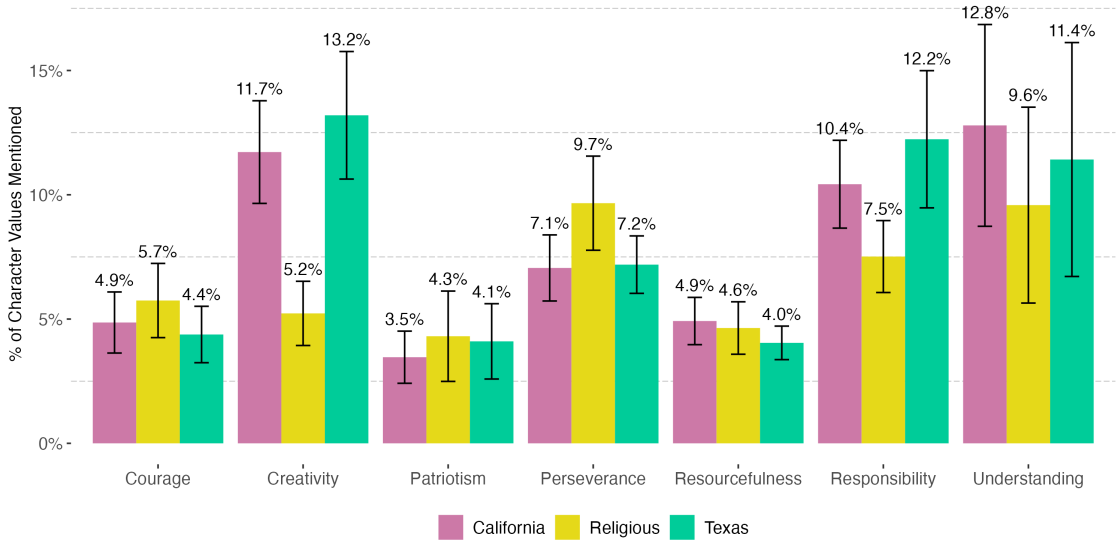
Notes: This figure shows the top values and topics (e.g. values and topics that appear on the highest average percentage of pages) within a collection. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection so that each subject-grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure A.5: Character Values

(a) Average Percent of Reading Pages Containing Specific Character Values Categories

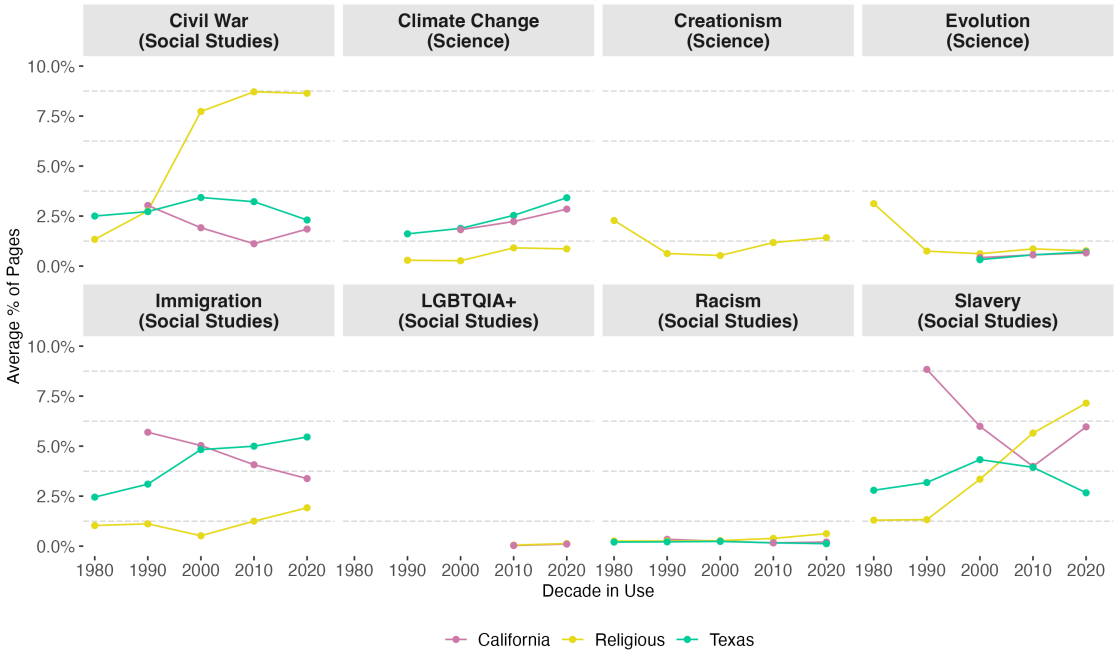


(b) Top Character Values by Collection



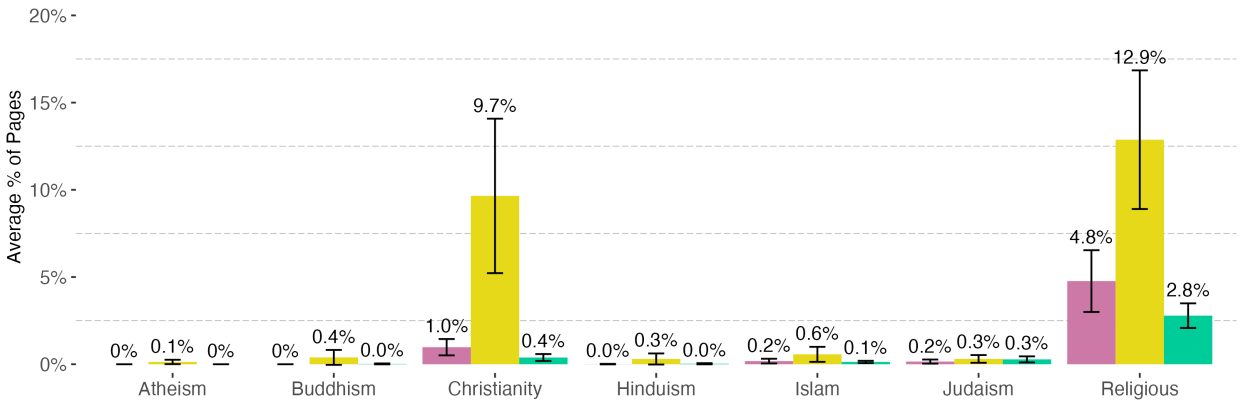
Notes: Panel (a) shows the average proportion of pages that contain character value groups from specific subcategories outlined in state standards as described in Appendix Table B.1. Panel (b) shows the average proportion of all character value mentions that is accounted for by the most frequently mentioned values in each collection by subject. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection so that each subject–grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals. Note that Panel (a) only examines individual character values while Panel (b) includes groups of character values (e.g. the Responsibility group contains values such as responsibility, responsibilities, citizen responsibility, taking responsibility, responsible behavior, etc).

Figure A.6: Average Percent of Pages Containing High-Profile Topics Over Time



Notes: This figure shows the average percentage of pages within a specific subject as denoted in the parentheses that contain a specific topic over time. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and subject so that each grade–year cell receives equal weight.

Figure A.7: Average Percent of Pages in Social Studies Textbooks Containing Religious Topics or Topics Related to a Specific Religion



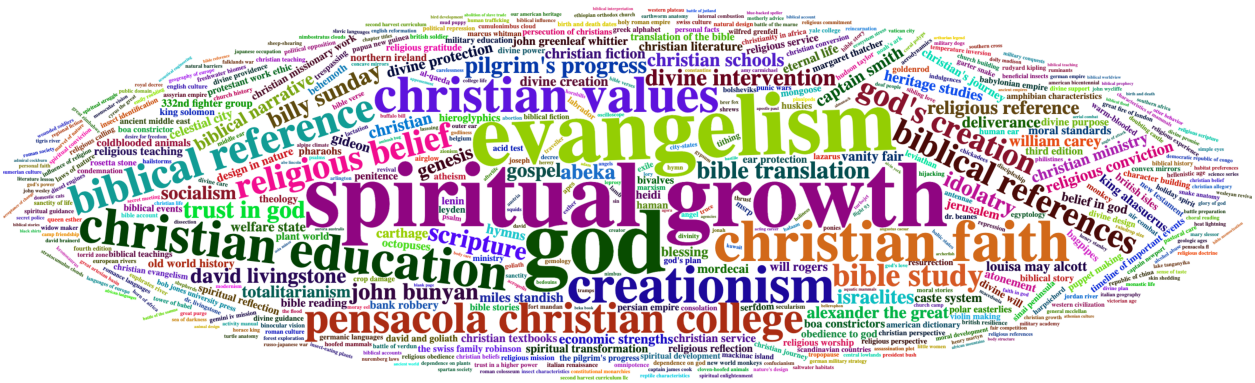
Notes: This figure shows the average percentage of pages within Social Studies textbooks that contain a religious topic or topics related to a specific religion. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and subject so that each grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure A.8: Topics and Values Exclusive to Each Collection: Word Clouds

(a) Topics and Values Exclusive to California Collection



(b) Topics and Values Exclusive to Religious Collection



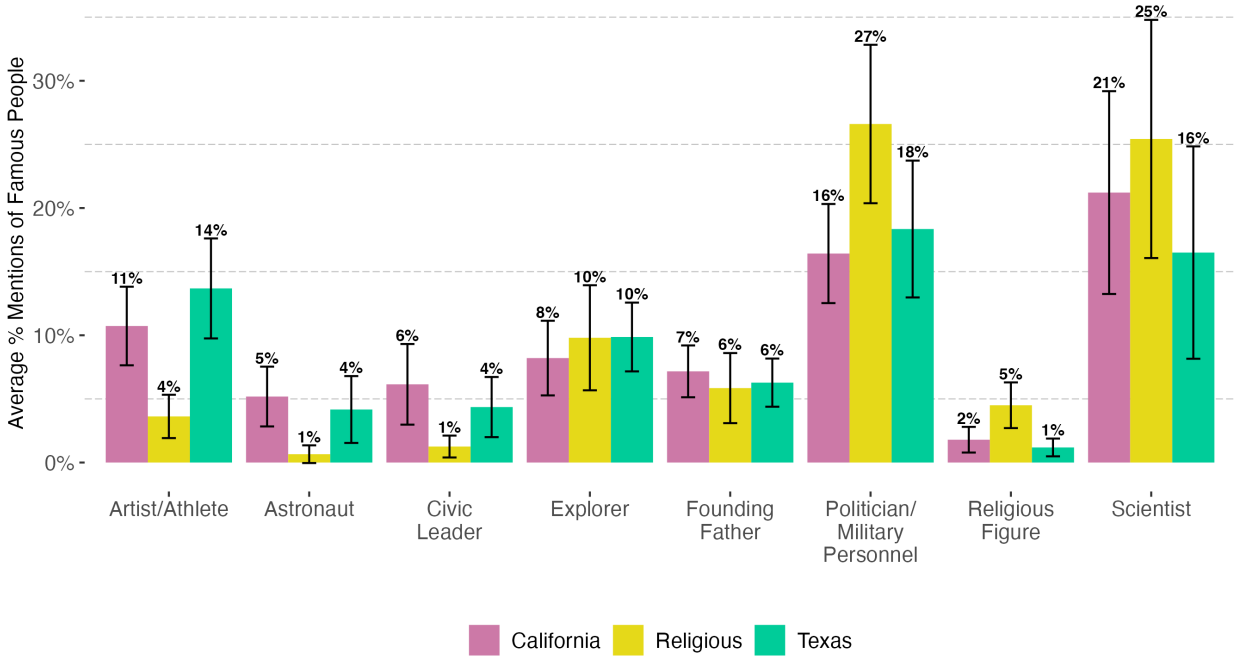
(c) Topics and Values Exclusive to Texas Collection



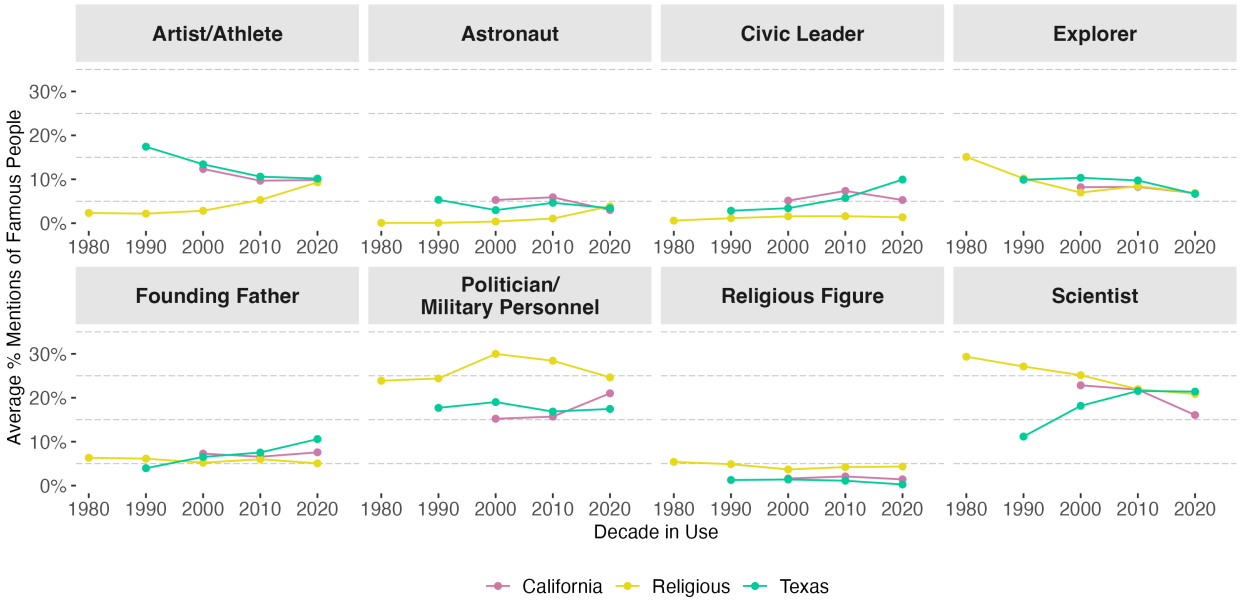
Notes: Each word cloud displays the topics and values metadata tags that appear exclusively within a single collection and not in either of the other two. Word size is proportional to frequency of occurrence across items in that collection. Publisher-specific tags (e.g., A Beka Book, BJU Press, Discovery Education) were excluded prior to generating the clouds.

Figure A.9: Historical Figures, by Occupation

(a) Overall



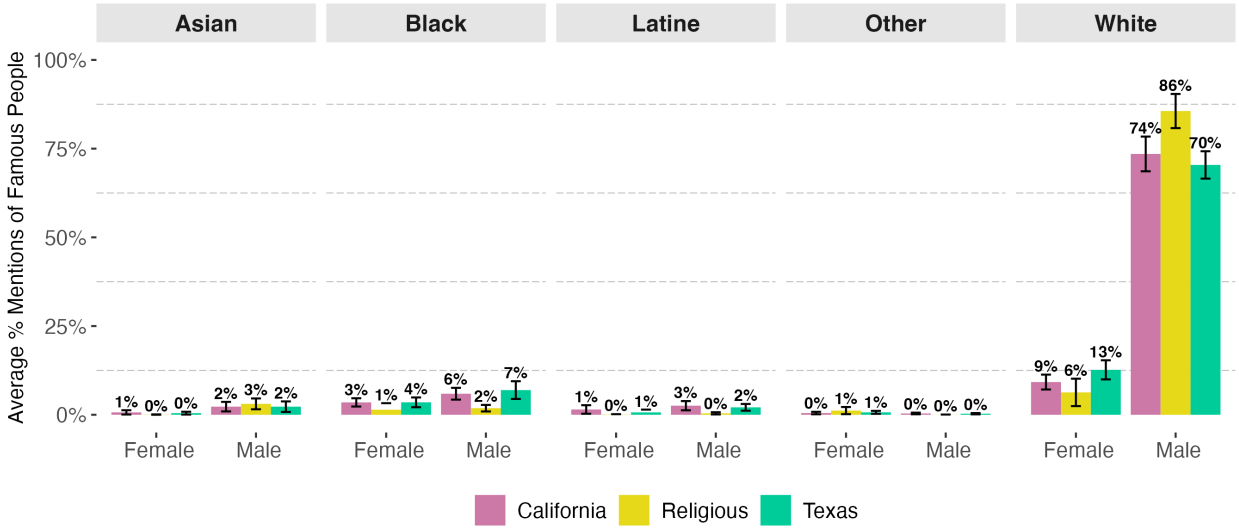
(b) Over Time



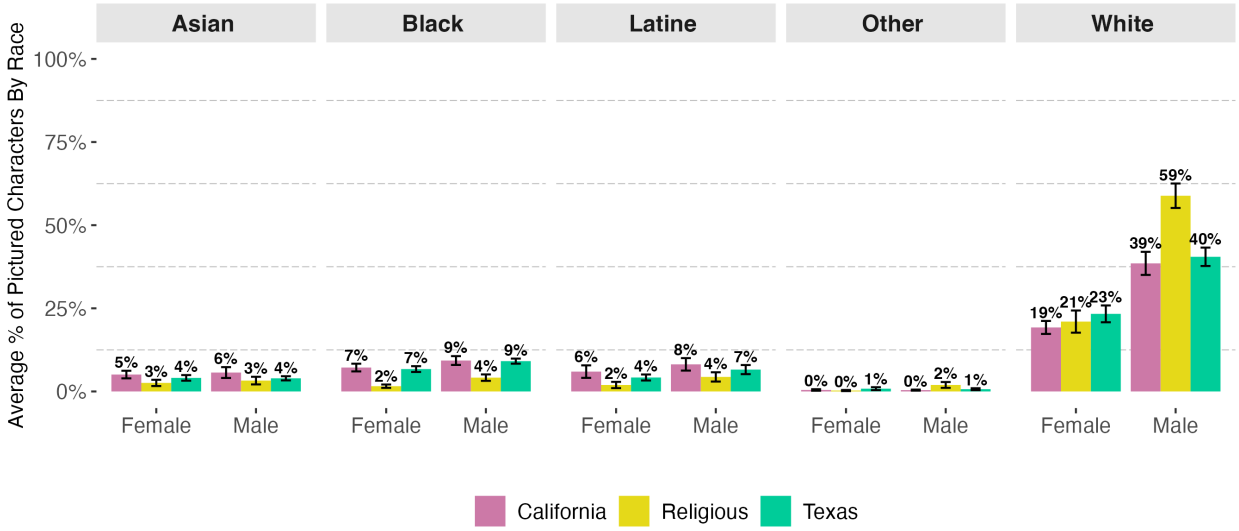
Notes: Panel (a) shows the average percentage of mentions of famous people by their occupation by collection and subject. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection and subject so that each grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals. Similarly, Panel (b) shows average percentage of mentions of famous people by their occupation by collection and decade in use.

Figure A.10: Race/Ethnicity and Gender Representation in Text and Images by Collection

(a) Text (Famous Individuals)



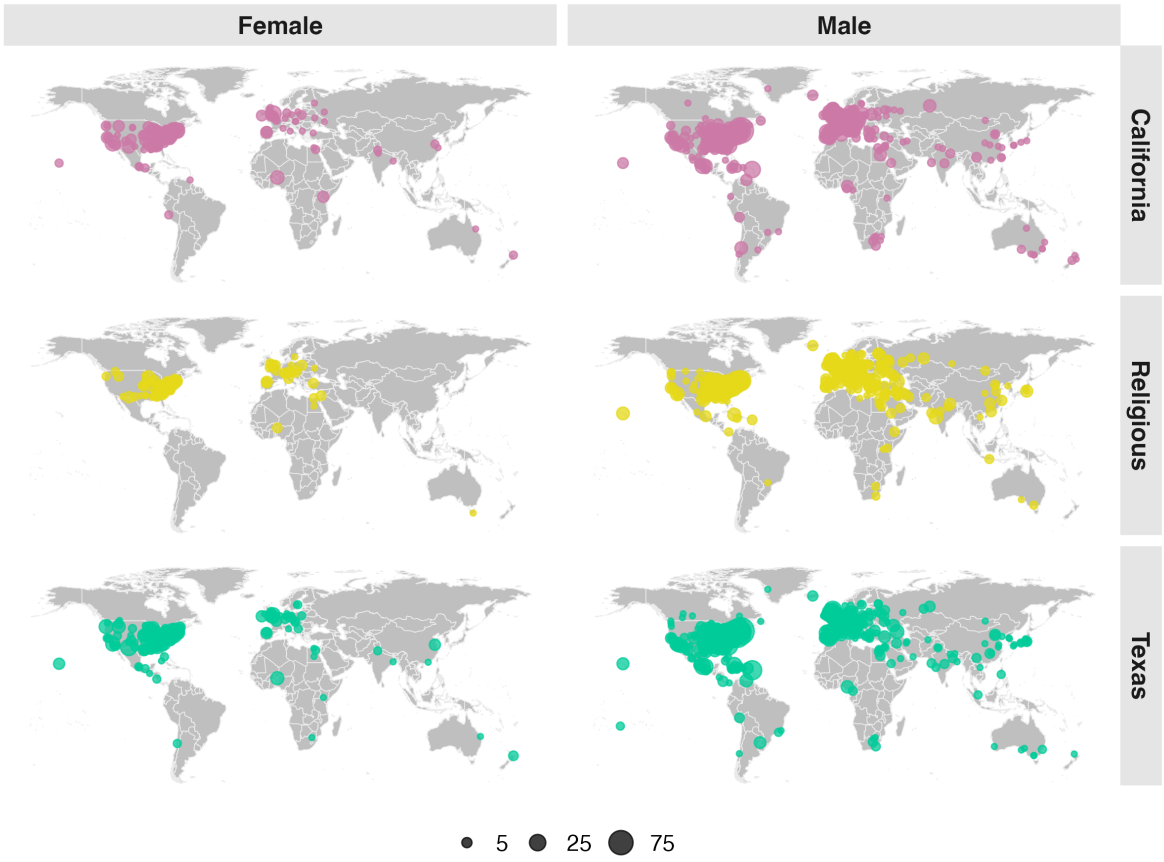
(b) Images



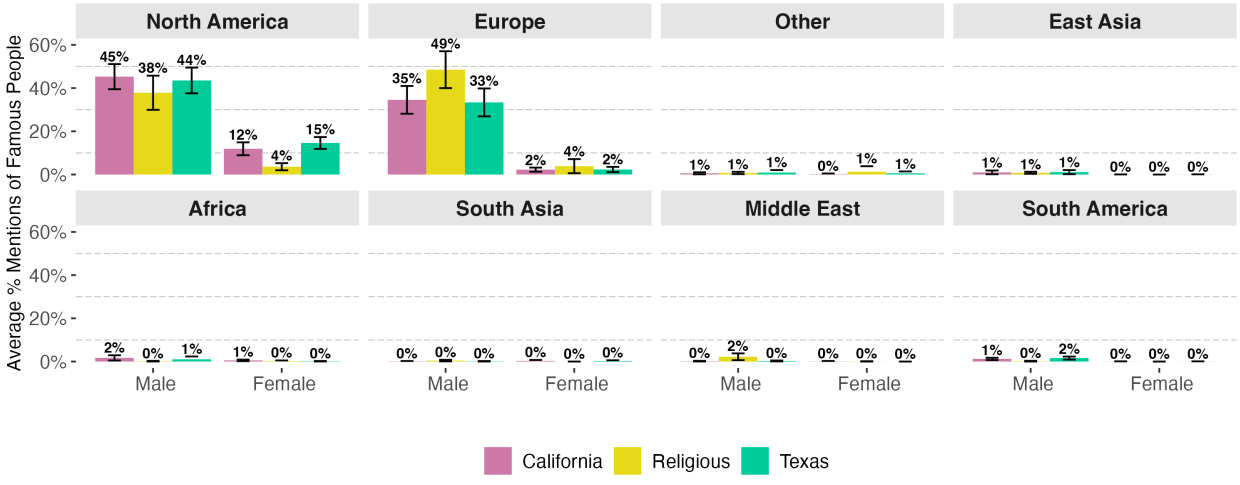
Notes: Panel (a) shows the average percentage of famous people identified as a specific race and gender out of all mentions of famous people by collection. Panel (b) shows the average percentage of pictured characters classified as a specific race and gender Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection so that each subject-grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure A.11: Birthplaces of Historical Figures by Gender

(a) Map



(b) Percent of Mentions



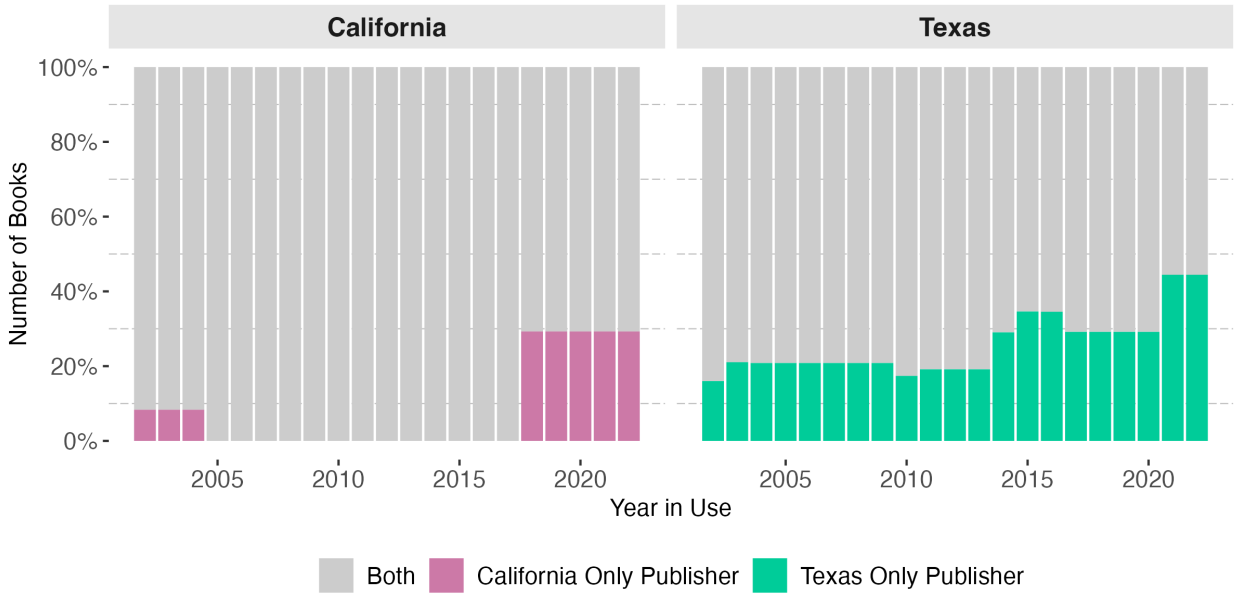
Notes: This figure shows the birthplaces of all famous people mentioned in a collection by gender. In Panel (a), the size of the point represents the number of mentions. In Panel (b) we show the average percentage of mentions for each birthplace by gender. Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection so that each subject–grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure A.12: Female Representation



Notes: This figure shows the average percentage of female words by collection over time (Panels (a) and (b)) and by collection and subject (Panels (c) and (d)). Averages are computed by first averaging textbook-level measures within collection × subject × grade × year-in-use cells and then averaging these cell means within each collection × decade cell (Panels (a) and (b)) or within each collection (Panels (c) and (d)) so that each subject–grade–year cell receives equal weight. Standard errors are estimated using weighted least squares and clustered at the textbook-level. Error bars denote 95% confidence intervals.

Figure A.13: Publisher Presence in Texas and California Textbook Adoptions



Notes: This figure displays the proportion of textbooks in use each year by publisher market presence, separately for Texas and California. Publishers are classified as "Texas Only" if they appear exclusively in Texas adoptions, "California Only" if they appear exclusively in California adoptions, and "Both" if they have books adopted in both states at any point in the sample. Data are drawn from all state textbook adoption records for Reading, Science, and Social Studies subjects in grades 3 and 5 from 2002–2022. Not that we include all adopted textbooks, not just the textbooks digitized in our sample. The top publishers are McGraw-Hill, Scott Foresman, Harcourt, Houghton Mifflin, and Pearson. Example of publishers with materials adopted by only Texas or only California include National Geographic Learning, Nystrom, Studies Weekly, Accelerate Learning, and Discovery Education

B Appendix Tables

Table B.1: Character Trait Emphasis Across Curriculum

	Texas	California	BJU	Abeka	A.C.E.
Trustworthiness	loyalty, integrity, reliability, punctuality	trustworthiness	—	truthful	trustworthy, truthful, faithful, loyal, honest, sincere, integrity, steadfast, consistent, committed
Responsibility	accountability, perseverance, diligence, self-control	responsibility	—	dependable, diligent, orderly, self-control	dependable, diligent, responsible, perseverant, punctual, self-controlled, prepared, thorough
Caring	kindness, empathy, charity, generosity, patience, consideration, compassion	caring	—	attentive, unselfish, patient, thoughtful, kind, gentle, generous, helpful	affectionate, kind, gentle, loving, compassionate, generous, friendly, thoughtful, considerate, helpful, gracious, forgiving, merciful, affectionate
Citizenship	respect, courtesy, concern for the common good and community, fairness, justice, patriotism, respect for authority and law, gratitude	citizenship	—	thankful, courteous	grateful, patriotic, tolerant, understanding, courteous, flexible, appreciative, tactful
Respect	(in Citizenship)	respect	—	respectful, polite	respectful
Fairness	(in Citizenship)	fairness	—	—	fair, equitable, just
Obedience	—	—	—	obedient	obedient, submissive, meek, orderly
Attitude	—	—	—	cheerful	joyful, peaceful, happy, content, rejoiceful
Religious	—	—	—	—	godly, holy, righteous, reverent
Other	—	—	—	—	discerning, repentant, modest, pure, zealous, virtuous, wise, prudent, creative, resourceful, observant, courageous, thrifty, purposeful, decisive, efficient, industrious, secure, temperate, humble

Note: Character traits emphasized in Texas curriculum are from the Texas Administrative Code, Title 19, Part 2, Chapter 120, §120.2 "Character Traits." California's *Six Pillars of Character* reflect the state's Department of Education guidance on character education posted on their website, based on the CHARACTER COUNTS! framework. Abeka traits are taken from its K5 "Character Development Visuals" lessons. A.C.E. traits are from the "90 Character Traits of Christ" included in A.C.E. curriculum materials.

Table B.2: Top 20 Famous People by Collection

Collection	Name	Race	Gender	Occupation	Sexuality	Rank	Mentions	Textbooks
California	Benjamin Franklin	White	Male	Founding Father	Hetero	1	383	32
California	Thomas Jefferson	White	Male	Founding Father	Hetero	2	322	31
California	George Washington	White	Male	Founding Father	Hetero	3	417	30
California	Abraham Lincoln	White	Male	Politician	Hetero	4	197	25
California	Martin Luther King Jr.	Black	Male	Civic Leader	Hetero	5	133	24
California	John Adams	White	Male	Founding Father	Hetero	6	233	18
California	Abigail Adams	White	Female	Companion	Hetero	7	81	18
California	Frederick Douglass	Black	Male	Civic Leader	Hetero	8	95	17
California	Paul Revere	White	Male	Military	Hetero	9	194	16
California	James Madison	White	Male	Founding Father	Hetero	10	139	16
California	Anne Hutchinson	White	Female	Religious Figure	Hetero	11	88	16
California	Susan B. Anthony	White	Female	Civic Leader	Hetero	12	83	16
California	John Hancock	White	Male	Politician	Hetero	13	67	16
California	Harriet Tubman	Black	Female	Writer	Hetero	14	64	16
California	William Clark	White	Male	Explorer	Hetero	15	57	16
California	Theodore Roosevelt	White	Male	Politician	Hetero	16	32	16
California	Alexander Hamilton	White	Male	Founding Father	Hetero	17	217	15
California	Neil Armstrong	White	Male	Astronaut	Hetero	18	145	15
California	George III	White	Male	Politician	Hetero	19	124	15
California	Patrick Henry	White	Male	Politician	Hetero	20	112	15
Religious	George Washington	White	Male	Founding Father	Hetero	1	310	33
Religious	Abraham Lincoln	White	Male	Politician	Hetero	2	369	27
Religious	Benjamin Franklin	White	Male	Founding Father	Hetero	3	158	25
Religious	Paul Revere	White	Male	Military	Hetero	4	185	19
Religious	Thomas Jefferson	White	Male	Founding Father	Hetero	5	146	18
Religious	Noah Webster	White	Male	Linguist	Hetero	6	280	17
Religious	John F. Kennedy	White	Male	Politician	Hetero	7	102	17
Religious	James Madison	White	Male	Founding Father	Hetero	8	50	16
Religious	Robert E. Lee	White	Male	Military	Hetero	9	132	15
Religious	Daniel Boone	White	Male	Explorer	Hetero	10	103	14
Religious	David Livingstone	White	Male	Explorer	Hetero	11	59	13
Religious	Ulysses S. Grant	White	Male	Politician	Hetero	12	56	12
Religious	Charles The Bald	White	Male	Politician	Hetero	13	45	12
Religious	William Bradford	White	Male	Politician	Hetero	14	31	12
Religious	Rudyard Kipling	White	Male	Writer	Hetero	15	20	12
Religious	John Ball	White	Male	Religious Figure	Hetero	16	12	12
Religious	John Smith	White	Male	Explorer	Hetero	17	123	11
Religious	Woodrow Wilson	White	Male	Politician	Hetero	18	31	11
Religious	Francis Scott Key	White	Male	Writer	Hetero	19	20	11
Religious	Hans C. Andersen	White	Male	Writer	Hetero	20	17	11
Texas	George Washington	White	Male	Founding Father	Hetero	1	531	45
Texas	Benjamin Franklin	White	Male	Founding Father	Hetero	2	377	33
Texas	Abraham Lincoln	White	Male	Politician	Hetero	3	422	31
Texas	Thomas Jefferson	White	Male	Founding Father	Hetero	4	245	31
Texas	Alexander Hamilton	White	Male	Founding Father	Hetero	5	245	27
Texas	Paul Revere	White	Male	Military	Hetero	6	268	25
Texas	Martin Luther King Jr.	Black	Male	Civic Leader	Hetero	7	219	24
Texas	Neil Armstrong	White	Male	Astronaut	Hetero	8	198	24
Texas	Harriet Tubman	Black	Female	Writer	Hetero	9	127	22
Texas	John F. Kennedy	White	Male	Politician	Hetero	10	214	21
Texas	John Adams	White	Male	Founding Father	Hetero	11	175	21
Texas	Theodore Roosevelt	White	Male	Politician	Hetero	12	157	21
Texas	Thomas Edison	White	Male	Inventor	Hetero	13	101	21
Texas	William Clark	White	Male	Explorer	Hetero	14	63	21
Texas	Francis Scott Key	White	Male	Writer	Hetero	15	58	21
Texas	Franklin D. Roosevelt	White	Male	Politician	Hetero	16	454	20
Texas	Langston Hughes	Black	Male	Writer	Hetero	17	68	20
Texas	Meriwether Lewis	White	Male	Explorer	Hetero	18	59	19
Texas	Lyndon B. Johnson	White	Male	Politician	Hetero	19	58	19
Texas	Jane Addams	White	Female	Civic Leader	Lesbian	20	92	18

Notes: This figure illustrates the most frequently famous people mentioned within a collection. We find the number of times a famous person is mentioned and the number of textbooks they appear in and then sort first by number of textbooks within a collection and then by number of mentions to find the top individuals.

Table B.3: Differences in Representation by Collection

	Dependent variable:				
	Skin Tint		% Female Words	% Female Characters	% Human Characters
	(1)	(2)	(3)	(4)	(5)
Religious	6.191***	4.865***	−0.095***	−0.072***	−0.037
	(1.343)	(1.378)	(0.023)	(0.024)	(0.030)
Texas	−0.356	−0.813	0.024	0.007	−0.018
	(1.675)	(1.704)	(0.020)	(0.016)	(0.025)
Asian		0.775			
		(0.881)			
Black		−13.333***			
		(0.676)			
Latine		−7.820***			
		(0.736)			
Other/Unsure Race		−1.311			
		(0.947)			
Grade 5	−4.915***	−4.872***	−0.062***	−0.012	0.033
	(1.290)	(1.302)	(0.022)	(0.027)	(0.025)
Science	−9.127***	−8.578***	−0.114***	0.048**	−0.247***
	(2.120)	(2.193)	(0.020)	(0.020)	(0.034)
Social Studies	−2.687*	−1.841	−0.209***	−0.108***	0.204***
	(1.442)	(1.467)	(0.019)	(0.020)	(0.018)
Year in Use	−0.108*	−0.109*	0.0003	0.0005	0.0004
	(0.059)	(0.060)	(0.001)	(0.001)	(0.001)
Constant	64.022***	66.688***	0.517***	0.388***	0.683***
	(2.305)	(2.317)	(0.032)	(0.028)	(0.038)
Observations	347,294	211,202	2,521	2,521	2,521
Adjusted R^2	0.084	0.162	0.339	0.164	0.427

Notes: Column 1 shows a regression of skin tint of a face detected in a given textbook on indicator variables indicating which collection, subject and grade the textbook belongs to. We also include the year a textbook was in use as a continuous variable centered around the first year a textbook was in use in our sample. In column 2, we also control for the predicted race of faces. The dependent variable in column 3 is the proportion of female words in given textbook. The dependent variables in columns 4-5 show the proportion of pictured human female characters and pictured human characters. Race, gender, and human/non-human predictions for pictured characters are from Gemini 2.5 Pro Preview as described in Section IV. Standard errors clustered at the textbook level. *p<0.1; **p<0.05; ***p<0.01

Table B.4: Top Relative Words by Gender

(a) Text

		Mean		
Top Words Associated with Female (vs. Male) Pronouns		V	A	D
California	mother, know, get, come, things, sure, way, couple, go, like	0.67	0.38	0.57
Religious	mother, mom, wife, get, know, daughter, crying, back, sure, tell	0.70	0.44	0.56
Texas	mother, get, know, come, things, sure, back, look, put, go	0.64	0.38	0.57
Top Words Associated with Male (vs. Female) Pronouns		V	A	D
California	cavalry, army, conquest, forces, territory, militia, troops, invasion, military, revolt	0.48	0.72	0.80
Religious	established, military, establish, army, forces, aim, major, notably, believed, history	0.61	0.56	0.75
Texas	army, forces, cavalry, military, territory, troops, armed, predecessor, war, militia	0.43	0.69	0.78

(a) Images

		Mean		
Top Words Associated with Female (vs. Male) Images		V	A	D
California	kids, classroom, children, gardening, decorating, healthy, coloring, homework, mom, sewing	0.76	0.41	0.47
Religious	mother, grandma, grandmother, aunt, mom, comforting, daughter, sister, nana, baby	0.84	0.32	0.52
Texas	mom, kids, children, grandma, homework, adult, kitchen, playroom, decorating, mother	0.80	0.43	0.52
Top Words Associated with Male (vs. Female) Images		V	A	D
California	cavalry, steamship, fort, steamboat, warship, regiment, garrison, lance, militia, republic	0.53	0.62	0.74
Religious	cavalry, regiment, army, militia, infantry, garrison, fort, troops, armored, lance	0.50	0.67	0.78
Texas	cavalry, militia, army, lance, fort, regiment, garrison, cannon, armored, war	0.47	0.68	0.77

Notes: This figure shows the top 5 relative words associated with a specific gender as well as the average Valence, Arousal, and Dominance (VAD) scores of those words. In panel (a), we use word embeddings estimated on the textbooks from each collection. In panel b, we use document embeddings estimated on descriptions of pictures males and females in the textbooks. To find the top relative words associated with females vs. males, we find the words that are closest to females in word embedding space and farthest from the males. We do the a similar exercise to find the top relative words associated with males vs. females. Note that we are limiting are analysis to words that are included in the VAD lexicon.

Table B.5: Famous Lesbian, Gay, Bisexual Individuals Mentioned by Collection

Collection	Name	Race	Gender	Occupation	Rank	Mentions	Textbooks
California	Jane Addams	White	Female	Civic Leader	1	24	4
California	Friedrich Wilhelm Von Steuben	White	Male	Military	2	4	4
California	Bessie Smith	Black	Female	Singer	3	5	3
California	Andy Warhol	White	Male	Artist	4	2	2
California	Michel Foucault	White	Male	Philosopher	5	5	1
California	Keith Haring	White	Male	Artist	6	3	1
California	Billie Jean King	White	Female	Tennis Player	7	2	1
California	Ellen Degeneres	White	Female	Comedian	8	2	1
California	Brian Boitano	White	Male	Skater	9	1	1
California	Claude Mckay	Black	Male	Writer	10	1	1
California	Frida Kahlo	Latina	Female	Painter	11	1	1
California	Harvey Milk	White	Male	Politician	12	1	1
California	James Baldwin	Black	Male	Writer	13	1	1
California	Ma Rainey	Black	Female	Musician	14	1	1
California	Tom Daley	White	Male	Athlete	15	1	1
Religious	James Baldwin	Black	Male	Writer	1	21	8
Religious	Elagabalus	Asian	Male	Politician	2	2	2
Religious	Pyotr Ilyich Tchaikovsky	White	Male	Composer	3	2	2
Religious	Simone De Beauvoir	White	Female	Writer	4	2	2
Religious	Jane Addams	White	Female	Civic Leader	5	2	1
Religious	Rupert Brooke	White	Male	Writer	6	1	1
Texas	Jane Addams	White	Female	Civic Leader	1	92	18
Texas	Barbara Jordan	Black	Female	Politician	2	13	3
Texas	Claude Mckay	Black	Male	Writer	3	9	3
Texas	Bessie Smith	Black	Female	Singer	4	5	3
Texas	Frida Kahlo	Latina	Female	Painter	5	4	2
Texas	Maurice Sendak	White	Male	Writer	6	2	2
Texas	Gertrude Stein	White	Female	Writer	7	3	1
Texas	Berenice Abbott	White	Female	Photographer	8	2	1
Texas	Michel Foucault	White	Male	Philosopher	9	2	1
Texas	Oscar Wilde	White	Male	Writer	10	2	1
Texas	Andrew Scott	White	Male	Actor	11	1	1
Texas	Billie Holiday	Black	Female	Singer	12	1	1
Texas	Greta Garbo	White	Female	Actor	13	1	1
Texas	Jason Brown	White	Male	Skater	14	1	1
Texas	Little Richard	Black	Male	Singer	15	1	1
Texas	Ma Rainey	Black	Female	Musician	16	1	1
Texas	Tom Daley	White	Male	Athlete	17	1	1

Notes: This figure illustrates the number of times a famous person identified as lesbian, gay, or bisexual (LGB) is mentioned within a collection. We calculate both the number of times a person as mentioned and the number of textbooks they appear in. We then sort first by number of textbooks within a collection and then by number of mentions. Our data on LGB famous people comes from the Pantheon 2.0 dataset. Note that this data does not contain information on transgender individuals.

C Methods Validation Appendix

In this appendix, we show the variety of ways in which we used manual validation to measure the effectiveness of our computational methods in measuring the people and topics present in our sample of educational curriculum.

C.A Feature Classification

In this section, we gauge the effectiveness of our feature classification model by comparing its performance with classifications completed by human annotators. The set of faces used in this validation procedure were randomly sampled from our textbooks and stratified by collection, subject and grade. Each randomly sampled face was randomly assigned 3 times to one of eight annotators. The annotators then determined whether the face was humanoid and, if so, labeled features of the face such as gender and race. Finally, the image was assigned the modal value out of the 3 labels, and any three way ties were labeled as "unsure." The annotators also had the option to mark their answers as "unsure" along with their response. For some features, we incorporate this uncertainty in the modal aggregation. This aligns with the feature classification model's structure, which could also predict "unsure" labels for certain features.

There were 1,080 images chosen for this exercise which were each labeled 3 times, resulting in 3,240 image-by-labeler observations. After subsetting these observations to those labeled as containing a human face, and then keeping only the images with 3 labelers, our final validation set for race and gender consists of approximately 900 images.

We show measures of model accuracy and inter-labeler reliability shown in Table C.1. To calculate the accuracy, we take the modal human response for each image as the true value and subsequently find the percentage of model predictions equal to said value. To calculate the inter-labeler reliability, we use the Fleiss' kappa measure commonly used to find the agreement between categorical labels given by a fixed number of labelers. This statistic ranges from 0 to 1 where 0 means complete disagreement and 1 means perfect agreement between labelers. In particular, we apply the following formula:

$$\hat{\kappa} = \frac{I_o - I_e}{1 - I_e} \tag{C.1}$$

where I_o is the observed agreement and I_e is the expected value based on chance.

Table C.1 shows that human agreement is consistently higher for gender than for race: without "unsure" labels, $\hat{\kappa} = 0.88$ for gender versus 0.65 for race, and when "unsure" is included $\hat{\kappa}$ declines to 0.72 for gender and more sharply to 0.45 for race. This indicates that race assignments are intrinsically more ambiguous and more sensitive to rater uncertainty in this labeling task. Model accuracy mirrors this pattern: both Gemini and FDAI achieve higher agreement with the modal human label on gender than on race, and both exhibit larger performance drops when "unsure" labels are retained, especially for race. Overall, the results suggest that lower race accuracy reflects, in part, lower underlying human consensus rather than purely model error, while gender labeling is comparatively stable for both humans and models. Conditional on the set of characters that have a face detected, gemini predictions consistently outperform the FDAI model predictions.

Table C.1: Accuracy and Inter-Labeler Reliability

(a) Without Unsure Labels

	Fleiss Kappa		Gemini		FDAI	
	$\hat{\kappa}$	N	Accuracy	N	Accuracy	N
Human	0.81	1,080	0.95	722	0.89	1,080
Race	0.65	602	0.79	415	0.57	602
Gender	0.88	779	0.81	536	0.65	779

(b) With Unsure Labels

	Fleiss Kappa		Gemini		FDAI	
	$\hat{\kappa}$	N	Accuracy	N	Accuracy	N
Human	0.81	1,080	0.95	722	0.89	1,080
Race	0.45	899	0.68	612	0.48	899
Gender	0.72	900	0.79	611	0.63	900

Notes: Accuracy is calculated as the percentage correct between the modal value of human labels and the predicted value for a given feature. Inter-labeler reliability was calculated using Fleiss' kappa, a measure of agreement between categorical labels assigned by a fixed number of labelers (here, 3). We repeat calculations with and without unsure labels. Race and Gender labels were only provided for faces labeled as Human.

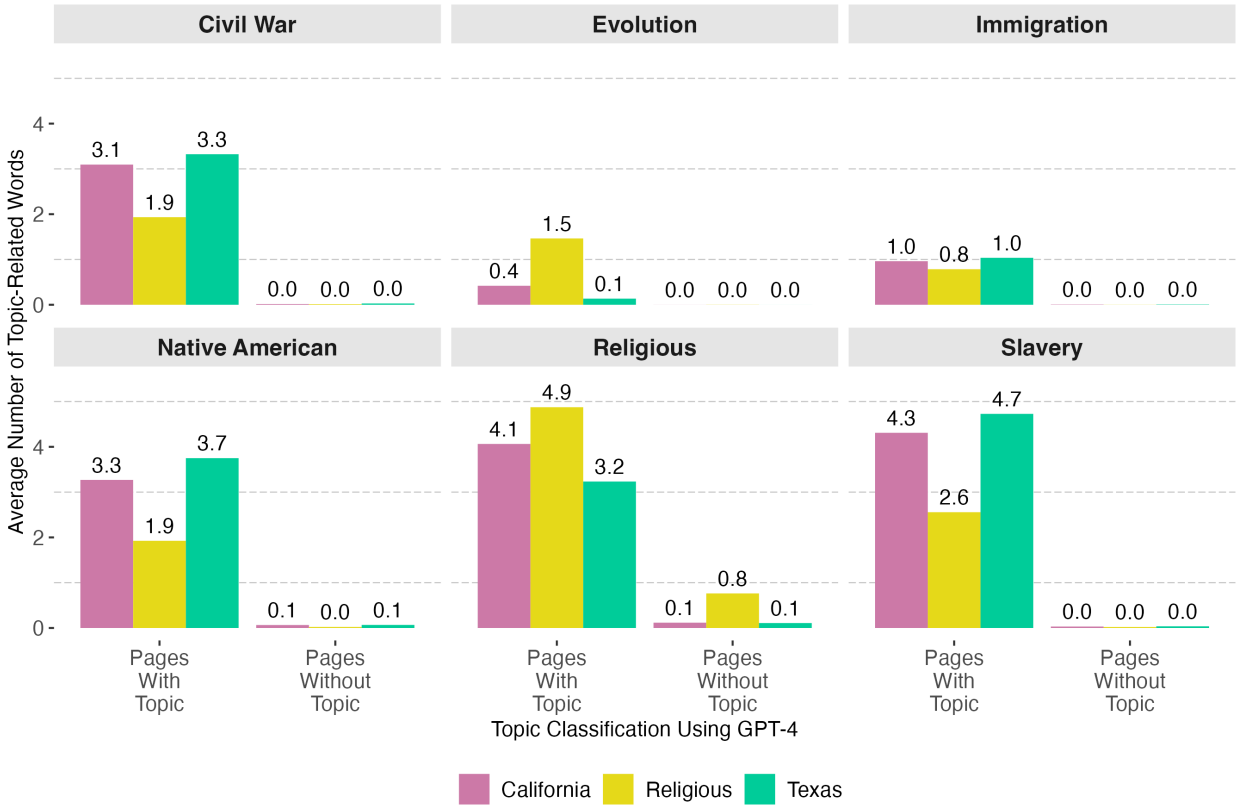
Note that for the gemini model this is only the accuracy of characters in our sample that **also** had a face detected and not the overall accuracy of the model. We merged characters detected and faces detecting using the largest overlap of bounding boxes. We were unable to match all of detected faces from FDAI to detected characters using this strategy which is why the sample sizes are smaller for our validation exercise for the gemini models. Our FDAI face detection model detected 59,769 faces and the gemini model detected 137,718 characters (over twice as many). This is in part because gemini finds more characters who do not have faces visible and more non-human characters. For this reason, the gemini characters that do not also have a face detected are 3.5 times more likely to be non-human, 4.5 times more likely to have an unsure gender label, and 6 times more likely to have an unsure race label. In our main analysis we drop gemini detected characters that have both an unsure race and an unsure gender prediction.

C.B Topic Classification

First we manually validate the topic classification performed by GPT using token counts of topic related words. In Figure C.1, we show that on average, pages that were not classified as containing specific topic, also do not have any mentions of those topic related words on the page. For example, pages that were not classified as containing topics related to "Slavery" have on average 0.0 words related to slavery such as captivity, enslavement, slave, plantation, Underground Railroad, Emancipation Proclamation, and Thirteenth Amendment. On pages that GPT classified as having the topic "Slavery," however, we see 3-4 words per page on average. A notable exception is the "religious/religion" topic. There

seem to be instances of words related to religion on pages that are not identified as having religious topics by GPT. Manual inspection of these pages yields examples such as a passage which discusses Martin Luther King, Junior and mentions that he was a pastor. On the corresponding page, there is a religious word, but religion is not a topic.[16]

Figure C.1: GPT-Identified Topics and Number of Topic-Related Tokens



Notes: This figure shows the average number of words on a page that are related to a given topic that are then also assigned to a predicted GPT topic classification. "TRUE" indicates that a given page has a topic-related word is related to a topic, and GPT also classified it as connected to that topic. "FALSE" indicates that the word is related a topic, but GPT did not classify it as such.

[16]Note that lexicon-based approaches may miss relevant terms and separately may include terms with multiple meanings that do not reflect the domain of interest.

D Methods Appendix

In this appendix, we discuss in further detail the application of some methods.

D.A Gemini 2.5 Pro Preview: Character Detection and Feature Classification

We detect characters in our textbooks using a two-stage, fully automated pipeline implemented using Google Cloud Vertex AI and Gemini 2.5 Pro Preview (model version `gemini-2.5-pro-preview-03-25`). In step one, we identify pages containing visual content and separate them from pages containing only text. For each scanned page, the raw image bytes are submitted to the model with a binary classification task: determine whether the page is `TEXT_ONLY` or `CONTAINS_IMAGE`. We configure this stage with a low temperature parameter (0.1) to minimize stochastic variation in model outputs.

In the second stage, we prompt the model to output a JSON object capturing: (i) scene-level data (what is happening or a setting description) and (ii) character-level data (bounding-box coordinates, race, gender, human or non-human, actions, location). We recode the model's race output into categories (Black, Latine, Asian, White, Unsure, Other, Not applicable). Additional post-processing steps harmonize labels and remove non-informative outputs.

D.B Image Description Embeddings (Pictured Human Characters)

To compare visual portrayals at scale, we construct embeddings from the LMM-generated text describing each pictured human character. We subset to human characters with non-missing character identifiers and construct one document per pictured character by concatenating the model's descriptions of `actions`, `location`, and `what_is_happening`. We define a unique identifier `page_character_id` by combining page and character identifiers.

Before training embeddings, we standardize multiword expressions using a curated bigram dictionary. Specifically, we map common alias phrases to canonical bigram tokens using a dictionary-based lookup procedure and then reconstruct the cleaned text. We further clean text for embeddings by enforcing UTF-8/ASCII characters, lowercasing, removing isolated single-letter OCR artifacts, normalizing apostrophes, removing numbers, removing punctuation (while retaining apostrophes), and collapsing repeated whitespace.

We then train a 300-dimensional Doc2Vec model using the PV-DM architecture (distributed memory). We set `vector_size=300`, `window=2`, `min_count=30`, and train for `epochs=400` with `hs=1` and `negative=0`; we set `workers=1` and a fixed random seed to support reproducibility. We initialize the model's word vectors using pretrained 300-dimensional embeddings and freeze these pretrained word vectors during Doc2Vec training. The resulting model produces one embedding per `page_character_id`, which we use for (i) book-level averages of character portrayals, (ii) top relative word and VAD analyses, and (iii) domain-association analyses.

D.C GPT-4 Preview: Topic and Value Classification

To extract the topics and values on a page we use OpenAI's API which allows us to define a specific LLM, a system role, and a *temperature* for each of our prompts. We use the `gpt-4-1106-preview` model. The parameter *temperature* defines the freedom or creativity we allow the model to have for its responses, and is set to 1. Finally, we assign a role to the

system to set its behavior. In our analysis, we instruct the model to assume the role of "a knowledgeable and unbiased judge". For each page we extract (i) a list of topics conveyed in the textbook page, containing no more than 10 topics, and (ii) a list of values conveyed in the textbook page, containing no more than 10 values.

D.D Skin Color

To extract and classify skin color, we begin by identifying the facial area of interest in each detected face by "segmenting" the skin. To do this, we use a face landmarker model,[17] which uses convolutional neural networks to predict and detect facial landmarks and then to isolate the parts of the face which contain skin in order to remove hair, clothing, and other accessories (known as "skin segmentation"). We then extract dominant colors by grouping pixels of the segmented skin into clusters using k-means clustering. We remove the smallest clusters, which tend to contain irrelevant information such as shadows or highlights, and treat the remaining cluster centroids as dominant skin tones. Finally, we collapse segmented skin into a single "representative skin color" by taking a weighted average of dominant skin colors in CIE L*a*b* color space (where weights are determined by cluster size). To construct a continuous measure of how dark or light the skin color of a detected face is, we use the L* value of this representative skin color which ranges from 0 (darkest) to 100 (lightest).

D.E ALC Word Embeddings (Text)

To characterize gendered context in our textbooks, we construct *à la carte* (ALC) embeddings for gendered pronouns using the `conText` framework. We begin with pre-trained 300-dimensional word vectors and estimate an ALC linear transformation matrix from our corpus of textbooks using a weighted word co-occurrence matrix (symmetric window size = 6, with weights that decay in distance). We then apply this transformation to the local contexts of gendered pronouns (*he*, *him*, *his*, *himself*, *she*, *her*, *hers*, and *herself*) to obtain one embedding for each pronoun occurrence. Because these pronoun-instance embeddings are estimated in the same semantic space as the pretrained word vectors, we can use cosine similarity to compare pronoun contexts to individual words and to broader semantic domains.

D.F Top Relative Words

To characterize the contexts most strongly associated with female versus male pronoun or pictured character instances, we identify "top relative words" using the pronoun-instance ALC embeddings or image description embeddings described above. This approach follows a similar methodology outlined in Manzini et al. (2019) and Charlesworth et al. (2022), but in our implementation the analysis is conducted at the instance level rather than using only one average embedding per group.

Let v_{ic}^{g} denote the embedding for instance i of gender g in collection c, and let $\bar{v}_{c}^{m}$ and $\bar{v}_{c}^{f}$ denote the average male and female embeddings within collection c. For a female instance in collection c, we define the relative query vector as:

$$q_{ic}^{f} = \text{norm}\left(v_{ic}^{f} - \bar{v}_{c}^{m}\right), \tag{D.2}$$

[17] We use Google's MediaPipe Face Landmarker.

and analogously, for a male instance:

$$q_{ic}^{m} = \text{norm}\left(v_{ic}^{m} - \bar{v}_{c}^{f}\right), \tag{D.3}$$

where norm$(\cdot)$ denotes length normalization. Intuitively, these relative query vectors capture the portion of an instance's contextual embedding that is distinctive relative to the opposite gender within the same collection.

For each instance, we identify the 50 nearest words to the relevant query vector using cosine similarity. To align this exercise with our sentiment analysis, we restrict candidate neighbors to words that appear in the NRC VAD lexicon (Mohammad, 2018). We then compute the mean valence, arousal, and dominance scores of these 50 nearest words and estimate collection-specific gender differences using regressions of the form

$$Y_{ibct}^{k} = \alpha + \sum_{c} \beta_{c} \left(Female_{i} \times \mathbf{1}\{Collection = c\}\right) + \lambda_{c} + \gamma_{t} + X_{b}'\delta + \varepsilon_{ibct}. \tag{D.4}$$

where Y_{ibct}^{k} is the mean score for dimension $k \in \{V, A, D\}$ for pronoun or character instance i in book b from collection c used in year t. $Female_{i}$ is an indicator for female pronouns or pictured characters. λ_{c} and γ_{t} denote collection and year-in-use fixed effects, and X_{b} includes controls for subject and grade. Standard errors are clustered at the book level. The coefficients β_{c} capture the collection-specific female-minus-male difference in the average VAD score of top relative words.

D.G Group-to-Domain Association

We also examine whether male and female contexts are more closely associated with specific semantic domains. In text, we look at instances of male and female gendered pronouns; in images we look at instances of male and female pictured characters. This analysis is related to the intuition behind the Word Embedding Association Test (Caliskan et al., 2017), but rather than estimating a WEAT test statistic, we compute an instance-level measure of similarity between each embedding and pre-specified domain word lists.

Let W_{D} denote the set of words belonging to domain D (for example, family, sports, tools, performance arts, home, or politics). For each instance embedding i, we measure association with domain D as the average cosine similarity between the embedding and the words in that domain:

$$s_{iD} = \frac{1}{|W_{D}|} \sum_{w \in W_{D}} \cos(v_{i}, w). \tag{D.5}$$

We then estimate collection-specific gender differences in domain association by estimating the following regression specification:

$$s_{iD} = \alpha_{D} + \sum_{c} \lambda_{Dc} \mathbf{1}\{C_{i} = c\} + \sum_{c} \beta_{Dc} \left(\mathbf{1}\{C_{i} = c\} \times Female_{i}\right) + X_{i}'\gamma_{D} + \varepsilon_{iDb}, \tag{D.6}$$

where X_{i} again includes controls for subject, grade, and year in use, and standard errors are

clustered at the book level. Positive values of β_{Dc} indicate that female contexts are more strongly associated than male contexts with domain D in collection c.

D.H Embedding Regressions

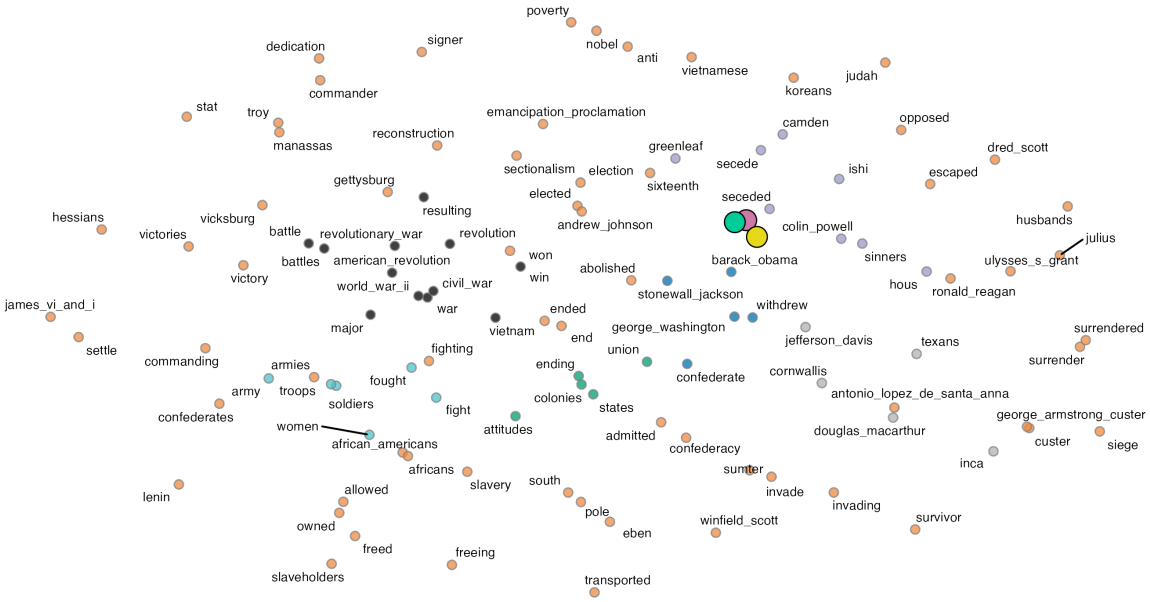
Following Rodriguez et al. (2023), we estimate $Y_i = X_i\beta + \varepsilon_i$ where Y_i is a 300-dimensional document embedding for book i, and X_i consists of j indicator variables which denote the collection, subject, grade, and copyright year of the book. The estimated coefficient vector $\hat{\beta}_j$ captures the average difference in embedding space between the reference group and the indicated group for a given dimension, holding other covariates constant. We summarize these vector-valued differences across dimensions by their Euclidean norm, $\|\hat{\beta}_j\|$, which measures the overall magnitude of semantic difference in embedding space between groups, aggregated across all embedding dimensions. Larger values indicate greater semantic differences between groups, while values near zero indicate little systematic separation in embedding space. We estimate two separate regressions; one where the Texas collection is the excluded group and one where the Religious collection is the excluded group. We report the normed coefficients from these regressions. These show the estimated differences across collections. Statistical significance is assessed via 1,000 permutation tests in which document embeddings are randomly reassigned across observations, the regression is re-estimated, and normed coefficients are recomputed. Two-sided p-values are calculated as the fraction of permuted normed coefficients that are greater than or equal to the observed norm.

D.I Survey Data

Our survey data come from the General Social Survey (GSS), a nationally representative repeated cross-sectional survey administered by NORC at the University of Chicago. Analyses use the restricted-use version of the GSS, which includes state FIPS codes back to 1975, enabling state-level comparisons not possible with the public release. Because not all variables of interest are available in every survey year, the year range varies by analysis; the specific waves used for each outcome are noted where relevant. To allow weights to contribute equally across years regardless of within-year sample size variation, wtssall was normalized within each survey year prior to pooling. To facilitate comparison across three groups — weekly churchgoers nationally, all residents of Texas, and all residents of California — respondents were duplicated across groups where applicable, with estimates computed separately within each group. Survey weights were applied consistently across all three groups using the survey package in R. These weights correct for unequal probability of selection arising from household size and other design features, and post-stratify to national age, sex, and education distributions. While the GSS is designed for national inference and its post-stratification targets do not correct for demographic composition within individual states, applying consistent weights across all groups ensures methodological comparability and preserves the within-household selection correction that remains valid at the state level. Analyses treat Texas and California estimates as weighted descriptions of GSS respondents in those states rather than fully representative state-level population estimates, and state-level results should be interpreted accordingly.

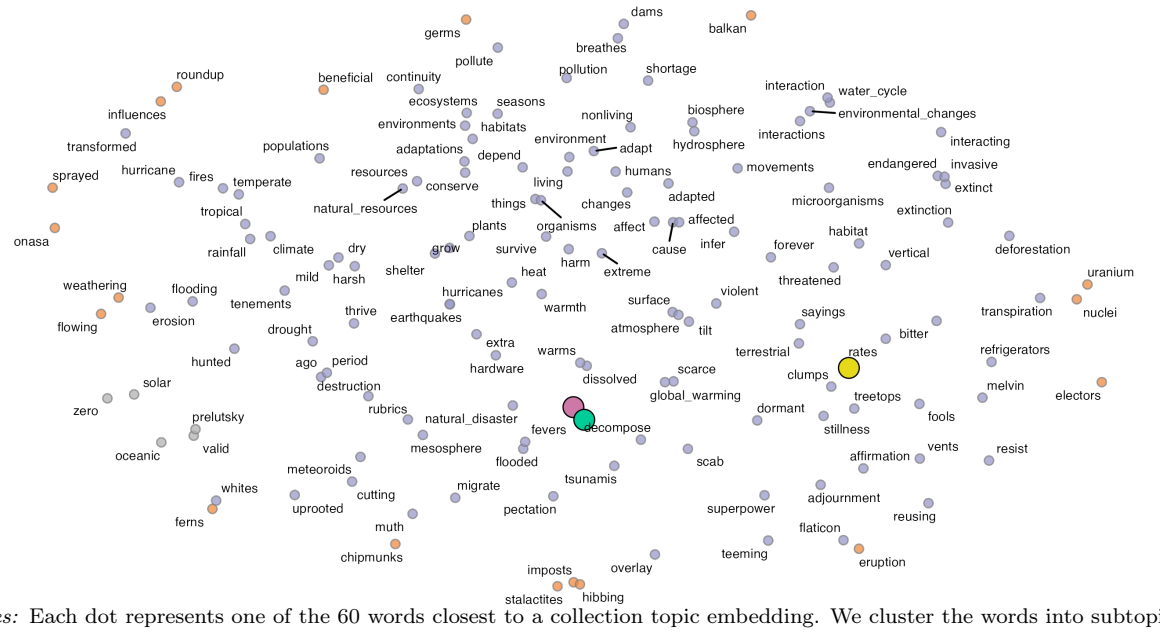
E Topic Embeddings

Figure E.2: Civil War: Topic Embeddings (Top Words)



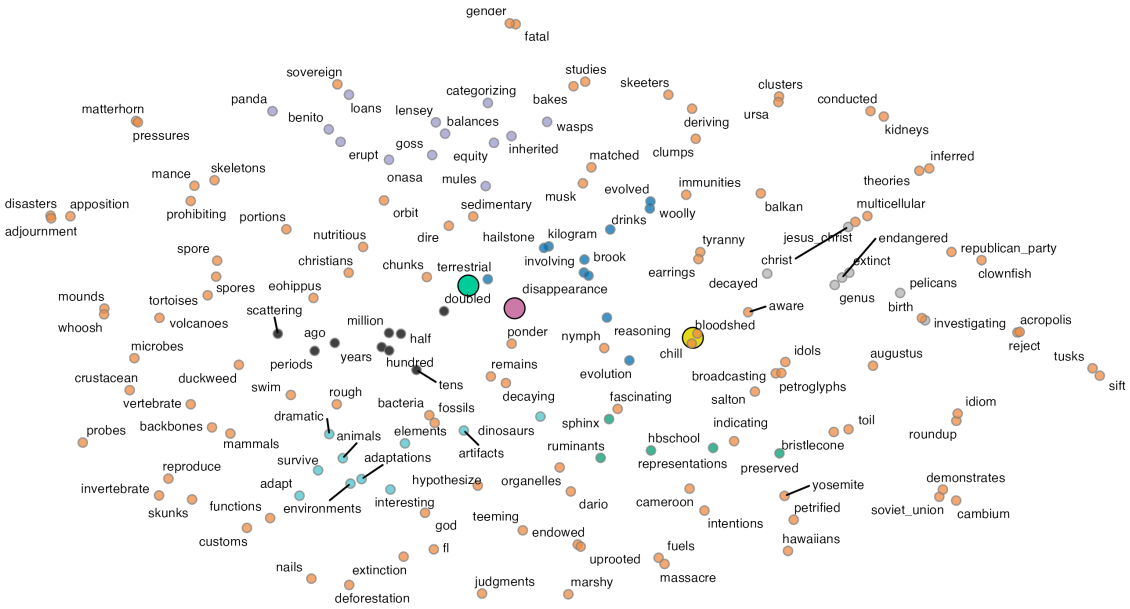
Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated is discussed in Section IV. These embeddings are reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embeddings observations should be interpreted in terms of their relative positions.

Figure E.3: Climate Change: Topic Embeddings (Top Words)



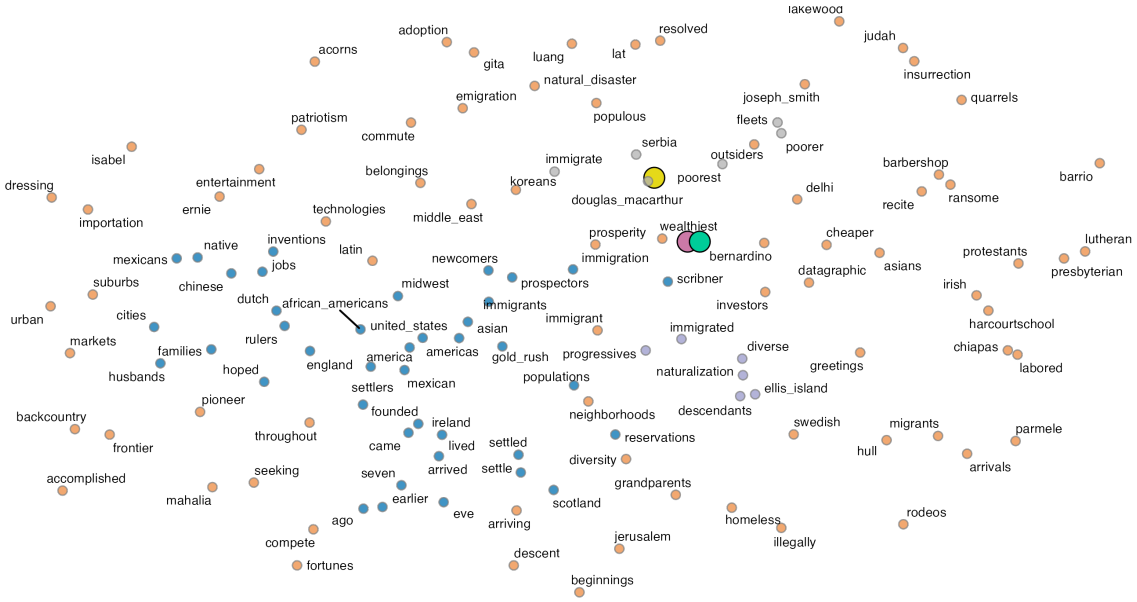
Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated is discussed in Section IV. These embeddings are reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embeddings observations should be interpreted in terms of their relative positions.

Figure E.4: Evolution: Topic Embeddings (Top Words)



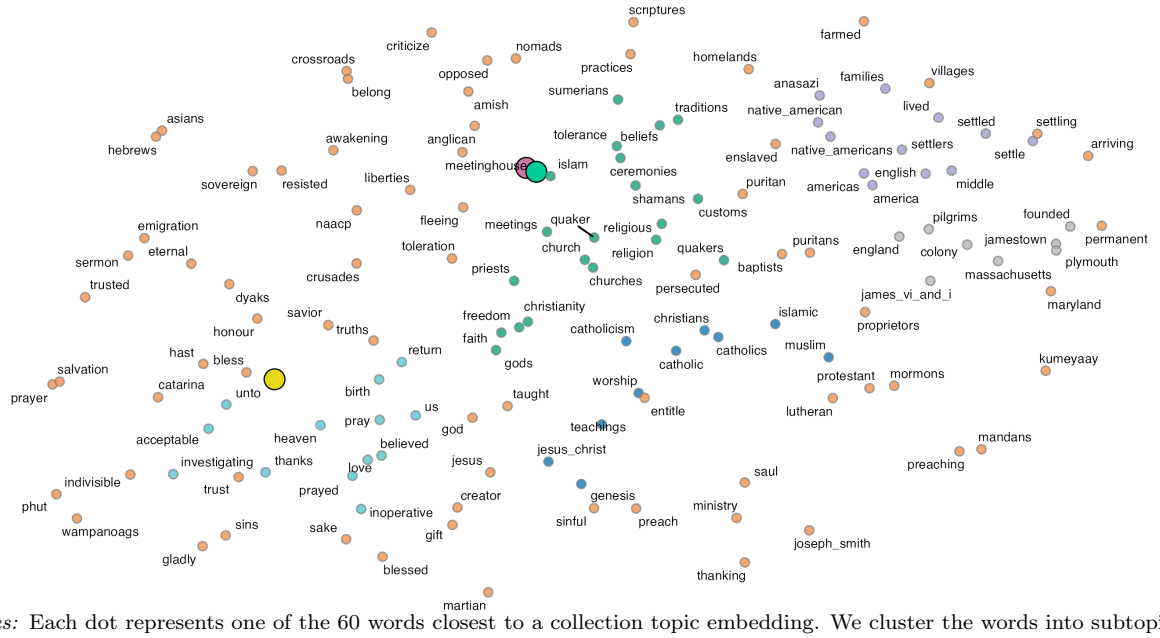
Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated is discussed in Section IV. These embeddings are reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embeddings observations should be interpreted in terms of their relative positions.

Figure E.5: Immigration: Topic Embeddings (Top Words)



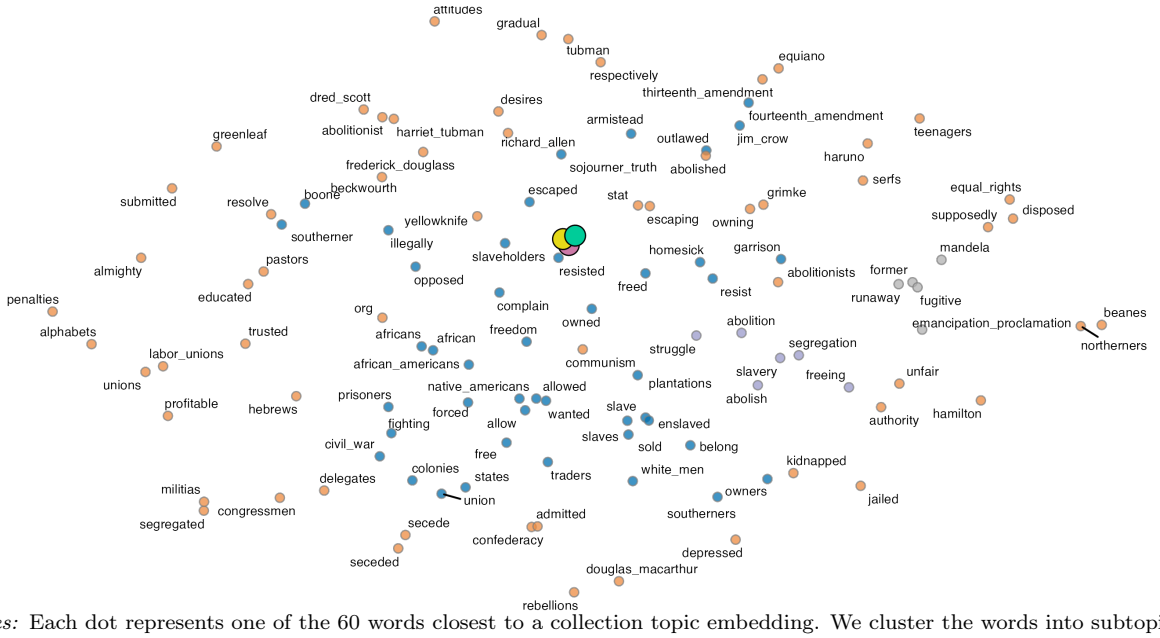
Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated is discussed in Section IV. These embeddings are reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embeddings observations should be interpreted in terms of their relative positions.

Figure E.6: Religious: Topic Embeddings (Top Words)



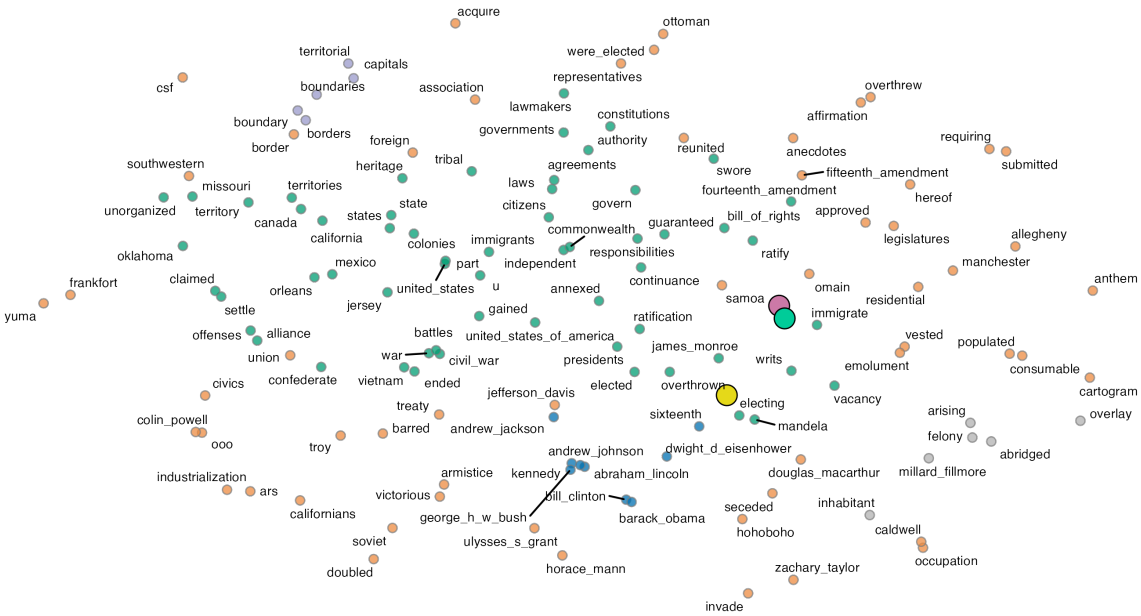
Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated is discussed in Section IV. These embeddings are reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embeddings observations should be interpreted in terms of their relative positions.

Figure E.7: Slavery: Topic Embeddings (Top Words)



Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated is discussed in Section IV. These embeddings are reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embeddings observations should be interpreted in terms of their relative positions.

Figure E.8: United States: Topic Embeddings (Top Words)



Notes: Each dot represents one of the 60 words closest to a collection topic embedding. We cluster the words into subtopics and denote these by color. Each large dot represents a collection topic embedding; pink for California, yellow for Religious, and green for Texas. Details on how these embeddings were estimated is discussed in Section IV. These embeddings are reduced to 2D via t-SNE while retaining meaningful information. Therefore, axis values do not have inherent meaning and embeddings observations should be interpreted in terms of their relative positions.