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 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, our findings reveal few meaningful differences between public school textbooks from Texas and California. However, religious school textbooks have a larger emphasis on character values such as caring and obedience, have less female representation, feature lighter-skinned individuals, and portray topics like evolution and religion differently. Important similarities also emerge: all textbook collections rarely include LGBTQIA+ discussion, portray females in more positive but less active or powerful contexts than males. Using public opinion survey data, we see 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 argue that the large overlap between Texas and California textbooks is consistent with the principle of minimal differentiation and find that differences in state content standards spill over 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, 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

I Introduction

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). 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. Given both the substantial and increasing allocation of tax dollars to these alternatives and the potential long-run consequences of educational content for children's beliefs and development, it is critical to understand what students are actually learning across different educational settings. 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 grade.

We apply computational social science methods to examine 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 have a great emphasis on critical thinking. Religious values are present in all three collections, however, religious values in public school textbooks are largely limited to the context of our nation's founding and religious values in our collection of religious textbooks 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); Ash et al. (2021); Adukia et al. (2022a,b, 2023); Voth and Yanagizawa-Drott (2024). For example, we

introduce a novel method to analyze the contexts in which characters are depicted in images by using generative AI tools to describe image settings and actions, and then using word embedding analyses to examine contextual bias in depictions across character identities. 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 II provides background and context related to school choice in the United States. Section III discusses prior textbook content analyses. Section IV describes the textbooks in our sample. Section V explains our image and text processing tools. Section VI presents our findings. Section VII describes the market and regulatory forces driving textbook content and Section VIII concludes. The appendices contain further information about the methods and additional exhibits presenting findings.

II 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, many parents 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 theory that individuals respond to perceived ideological threat by taking actions to preserve and transmit their preferred beliefs to their children (Wheaton, 2020). However, evidence on the academic benefits of different 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 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

[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).

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.

III 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

[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).

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.

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) who examine children's books, Ash et al. (2021) who examine newspapers, Voth and

Yanagizawa-Drott (2024) who examine yearbooks, and Kamel and Woo-Mora (2023) who examine sports statistics.

IV 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]

IV.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 materials 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 adopted list and districts generally ordered from the bulletin through the state system and the state paid publishers directly. Consequently, the adopted list provides a reasonable proxy for the textbooks in use. In 2011, Texas shifted to a more flexible funding and procurement regime (the Instructional Materials Allotment), under which districts could use state funds to purchase state-adopted materials or other instructional materials obtained directly from vendors, subject to annual certification that materials cover the Texas Essential Knowledge and Skills (TEKS), which are the state curriculum standards. In light of this, we first digitized a con-

[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.

venience 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 three textbooks purchased by school districts each year for each relevant grade and subject (reliable textbook purchasing data were not collected before 2011).

IV.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.

IV.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 school 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 tax-payer funding (Klein, 2017). We focus on textbooks published by these three Christian publishers both because of their wide adoption, but also because there are many taxpayer dollars funding schools which use these curricula.[5]

Using Google Trends search data as a measure of interest in these publishers (Figure 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

[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.

[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 2016-2017 figures from the Indiana Department of Education (Klein, 2017).

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). The religious publishers in our sample conveniently provide such options for families; one can simply go online and purchase an entire grade's worth of curricular materials directly from one of these publishers or from second-hand resellers.

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. 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..

V 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.

V.A Images As Data

The process that we use to convert images into data is illustrated in Figure 2a.

V.A.1 *Images as Data: Character Detection and Face Detection*

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 by harnessing the LMM's capability to recognize facial expressions, body language, and other visual cues, enabling us to extract detailed and nuanced information about the pictured characters. Importantly, unlike typical face detection models, this allows us to detect characters whose faces are not visible, as well 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. Additional details provided in our Methods Appendix D.

We supplement this character detection with a convolutional neural network (CNN)

classifier that detects faces using Google's Vertex AI platform. We use a face detection model trained to better detect faces in both photographs and illustrations as described in Szasz et al. (2022) and Adukia et al. (2023).

V.A.2 *Images as Data: Feature Classification of Putative Race, Gender*

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 that we trained using a manually curated data set of both illustrations and photographs which can more accurately predict features of illustrated faces such as the race and gender, as described in 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 race and gender predictions of this CNN model by a wide margin. This and other validation exercises are discussed in our Validation Appendix C.

V.A.3 *Images as Data: 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,[6] 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"). Then, we extract the dominant colors by grouping pixels of the segmented skin (based on their position in RGB color space) into clusters using k-means clustering. This groups each pixel into k clusters that share the most similar RGB color space, and each cluster is represented by the RGB value of its center. We remove the smallest two clusters, which tend to contain irrelevant information such as shadows or highlights. We then select the centers (or centroids) of the remaining clusters–which represent the most common skin-like colors–are selected. These centroids are then treated as the dominant skin tones of the detected face. Finally, we collapse the remaining segmented skin into a single "representative skin color" by taking a weighted average of the dominant skin colors in CIE L*a*b* color space (where the weights are determined by the size of the cluster from which each dominant color was drawn). 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) as our continuous measure of skin tint that serves as our measure of skin color. This process follows a similar method as described in Adukia et al. (2023) and is illustrated in Figure 2a.

[6]Specifically, we use Google's MediaPipe Face Landmarker.

V.A.4 *Images as Data: 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. We then train a 300-dimensional Doc2Vec model (PV-DM) on these character-documents. We use these embeddings in three ways. First, we construct a textbook-level embedding by averaging the (normalized) character-description embeddings within each textbook and compare these book-level embeddings across collections to summarize overall differences in how characters are portrayed. Second, 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. Finally, we measure the similarity between image descriptions and pre-specified semantic domains such as family, sports, home, and politics. More details can be found in Appendix D.

V.B Text As Data

The process that we use to convert text into data is presented in Figure 2b.

V.B.1 *Text as Data: Large Language Models*

We measure the topics and values included in the text using a large language models (LLMs). We use OpenAI's GPT-4 model (GPT), to analyze single pages of textbooks and identify topics and values present in them. More details about our prompt and parameter choice in Appendix Section D. We validate the model's topic classification by counting how often a set of topic-related words appear on pages that were (or were not) classified as containing that topic by GPT. We find that pages which were classified as having a specific topic were much more likely to contain words related to that topic (Figure C.1). More details about this, and other validation exercises in Appendix Section C. The described validation method presents another approach to measure the prevalence of specific topics, i.e., measuring counts of specific keywords instead of relying on a large language model to summarize the text. This would involve providing a list of topics as well as a comprehensive list of keywords for each subject. We use the GPT classifications to help identify topics, which helps us identify prevalent topics that we might not have thought to look for and systematically test for differences between the prevalence of topics across collections. This application of GPT is one example of using machine learning as a tool for hypothesis generation as discussed by Ludwig and Mullainathan (2024).

V.B.2 *Text as Data: Named Entity Recognition*

Named Entity Recognition (NER) identifies "named entities" (nouns predicted to be proper names such as locations, events, or people) which allows us to count their occurrences in the text. This provides measures of both what topics and which identities appear, including geography, historical events, and the race, gender, occupation, birthplace, and sexuality of famous individuals. To measure geographic exposure, we use the "GPE" (geopolitical entities) 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, which contains information on over 70,000 Wikipedia biographies, allowing us to identify figures such as Martin Luther King Jr. and Amelia Earhart (Yu et al., 2016). Using these data, we examine mentions of famous individuals by gender, occupation, sexuality,[7] and birthplace, and we manually classify their putative race. We also aggregate certain occupations into broader categories of interest.[8] Finally, we use remaining "PERSON" entities that are not matched to Pantheon to measure mentions of non-famous individuals. We extract first names and estimate the probability that each name is female or male using Social Security Administration data on name frequencies by gender.

V.B.3 *Text as Data: Token Counts*

Token counts are one of the simplest natural language processing tools which allow us to count the number of times specific groups or topics are mentioned. We can use these counts to see how often words related to females (such as *she*, *queen*, and *sister*) are mentioned relative to words related to males (such as *he*, *king*, and *brother*. These words include pronouns and other gendered terms.

V.B.4 *Text as Data: Sentiment Analysis using VAD*

To characterize the sentiment of words related to and the actions taken by agents of different identities, we measure three attributes of the words: their valence (V), arousal (A), and dominance (D). Valence measures range from happiness to unhappiness, or positive to negative. Arousal measures indicate a level of "affective activation," or active to passive. Dominance measures reflect the level of control of the emotional state, or dominant/powerful to submissive/weak. We use the NRC VAD Lexicon which includes a list of more than 20,000 English words and their valence, arousal, and dominance scores (Mohammad, 2018). Scores range from zero to one. For example, the words "teamwork" and "leader" have {V, A, D}

[7] If sexuality is missing for a famous person, we impute "Heterosexual."

[8] For example, we separately examine the presence and portrayal of Founding Fathers following Morris (1973), including John Adams, Benjamin Franklin, Alexander Hamilton, John Jay, Thomas Jefferson, James Madison, and George Washington. Broader definitions include all signers of the Declaration of Independence, Articles of Confederation, or U.S. Constitution, so our measure represents a lower bound.

scores of {0.92, 0.49, 0.77} and {0.83, 0.58, 0.93}, respectively. Both have relatively positive valence, but "teamwork" has slightly more positive association to it. However, "leader" has a higher arousal score and a meaningfully larger dominance score than "teamwork."

V.B.5 *Text as Data: Embeddings*

We use embeddings to characterize the semantic contexts in which different groups, topics, and documents appear in the textbook corpus.

ALC Word Embeddings. 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[9]. 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. We use these embeddings in two ways. First, we identify the words most distinctively associated with female versus male pronoun contexts and examine the valence, arousal, and dominance of those words. Second, we measure the similarity between pronoun contexts and pre-specified semantic domains such as family, sports, home, and politics. More details can be found in Appendix D.

Document Embeddings (Book and Page Embeddings). In order to characterize larger sections of text such as at the page or book level, we apply an analogous approach to word embeddings called "document embeddings." Similar to word embeddings, document embeddings are high-dimensional vector representations of documents. Unlike word embeddings, document embeddings can capture the representations of input texts of variable length such as sentences, paragraphs, pages, or entire books as opposed to single words. We define a document to be either an entire textbook or a single page in a textbook. We train separate Doc2Vec models using the distributed memory (DM) architecture from which we estimate a set of textbook embeddings (one embedding per *textbook*), and separately, a set of page embeddings (one embedding per textbook *page*). In a similar manner to how we analyze word embeddings, we use cosine similarity to calculate how "close," or similar, textbooks in our sample are to each other across collections, subjects, and time.

To visually assess similarities between textbooks, we need to apply techniques that reduce the dimensionality of these vectors to two-dimensional space without losing the de-

[9]We use symmetric window size = 6, with weights that decay in distance

tailed information that larger vector spaces provide (such as 300-dimensional space). To do so, we implement t-SNE to reduce these 300-dimensional vectors into a two-dimensional space. There are other approaches one could use (e.g. PCA), but t-SNE is better suited for this task, since it is more robust to outliers, less sensitive to the ordering of the data points, and better at preserving the clusters in the data.

Topic Embeddings. Finally, we integrate the output from GPT, which classifies topics on each page, with the page embeddings. We use this information in two analyses. First, for each topic, we calculate and average the pairwise cosine similarity of pages belonging to that topic for each pair of collections. This allows us to see, in aggregate, the semantic similarity of topics between collections. In the second analysis, we create a topic embedding for each topic and collection. To do so, we compute a single embedding by averaging the normalized page embeddings of all topic-related pages in a given collection. Hence, there will be 3 topic embeddings for each topic, one for each collection. Then, using the word embeddings which were estimated in the same embedding space, we apply cosine similarity to find the closest words to each topic embedding. This allows us to describe, with greater granularity, the context in which topics are discussed in each collection based on a visual examination of proximity between words. We also use embedding regressions as described in Appendix Section D to formally test for differences between topics.

VI Results

In this section, we discuss the content patterns that we uncover in our collections of textbooks. We begin by examining the *overall similarity* across the collections, by subject and over time.

VI.A Overall Patterns in Textbooks

First, we analyze the overall similarity between the text and images textbooks by using document embeddings to measure distances between textbooks in embedding space. The closer two textbooks are positioned in embedding space, the more similar their content is, both in terms of the words used and the context in which they are used.

We begin with a descriptive view of the embedding space. Figures 3a and 3b 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.[10] 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

[10]For visualization, we project the 300-dimensional embeddings into two dimensions using t-SNE; the figures should be interpreted in terms of books' relative positions in the reduced space.

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 3c and 3d, which provide measures of the overall magnitude of semantic difference in embedding space between groups, aggregated across all embedding dimensions. These regression 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. Once subject is held fixed, the principal collection-level separation in embedding space is between public-school and Religious textbooks rather than between California and Texas.

Figure A.2 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.2a distribution of cosine similarities between Texas and California textbook pairs is skewed toward higher similarity, compared to the distributions for California versus religious school textbooks or Texas versus religious school textbooks. Notably, some textbook pairs have near-perfect similarity scores, consistent with publishers reusing closely related content across state editions (See Section VII for a discussion of the textbook publishing process). Averaging these similarities over time yields the same ranking in every decade: California and Texas are always more similar to each other than either is to the Religious collection (Figure A.2b). However, in the most recent decade, the similarity between Reading textbooks from Texas and California has decreased, reaching levels of similarity (or dissimilarity) comparable to those with textbooks from our Religious collection. Initially, Science textbooks across all collections exhibited the highest similarity, but this has declined over time. Texas Science textbooks are generally more similar to California Science textbooks than to those from religious schools; over the past three decades, however, Texas Science textbooks have become less similar to those from California and more similar to religious school textbooks. By contrast, California and Texas Social Studies textbooks have grown substantially more similar to each other, while religious school Social Studies textbooks have become slightly more aligned with public school textbooks in the most recent decade.

The next sections move from these aggregate embedding patterns to specific content dimensions, which allows us to identify on where California and Texas still differ and where the Religious collection departs most sharply from both. We start with a discussion of the

values present in each textbook collection.

VI.B Values

We begin by exploring the presence of values related to guiding principles that may drive beliefs or attitudes and influence behavior and decision-making. The top values we identify are similar across collections with the top three values being centered around learning. We see this overlap especially between California and Texas textbooks (Figure A.3). The primary outlier is that of *faith*, which shows up as a top value in the Religious collection (Figures A.3 and 4a); this is consistent with the large proportion of religious values found in the religious collection.

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). Each category has a detailed set of traits listed (e.g., perseverance, respect for authority, patriotism). In contrast, California provides nonbinding guidance, recommending a six-part framework: trustworthiness, respect, responsibility, fairness, caring, and citizenship (California Department of Education, n.d.). These *Six Pillars of Character* align with national models like CHARACTER COUNTS! These mirror the four Texas pillars, though Texas bundles fairness and respect into citizenship. 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 private school curriculum 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. We will 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 Appendix Table B.1.

In Figure 4a, we see that the religious collection has a larger proportion of pages that containing character values than Texas and California textbooks. We dig in deeper to see which specific set of character values are driving these differences, and whether Texas and California differ on the emphasis they put on specific character values, particularly given the legal vs. suggested mandates across states. We see that across each of the character value categories, there are no statistically significant differences between Texas and California (Figure 4b). Similarly, there are no significant differences between the public-school and Religious collections in the presence of character traits such as those related to attitude, citizenship, and those other traits highlighted by A.C.E. However, we see that the difference in character values in Reading textbooks from our Religious collection is largely driven by character values related to caring, responsibility, and trustworthiness.

Religious Values Given that the Religious collection is 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 are present on 5% percent of pages on average in our collection of religious school textbooks as compared to 0.8%-2% of pages in our collections of public school textbooks (Figure A.4).

VI.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. Parents choose schooling options, in part, because they want their children to be exposed (or not) to certain knowledge. In this section, we document the topics emphasized in each of our textbook collections.

We first note the overall top topics that are addressed by subject (Figure A.5). We see that similar topics are discussed across collections within subjects with some notable exceptions unique to each collection. In Reading textbooks, two of the top three topics identified were the same for each collection: family and nature, with education being a top topic for both of the public school collections. The Religious collection is more likely to include narratives related to family than the collections of public school textbooks (16% vs. 10% or 12%) (Figure A.5a). It is also more likely to have narratives related to survival, prayer, and faith. In Science textbooks, astronomy, biology, physics, observation, and science are among the most common topics across all three collections (Figure A.5b). In Social Studies textbooks, geography, history (including American history), and agriculture are among the most common topics (Figure A.5c). Note that we see religion as a common topic in Social

Studies textbooks for our collection of religious school textbooks but not in the public school textbooks. Moreover, slavery is a common topic in California textbooks, whereas the Civil War is a common topic in the Religious collection.

High-Profile Topics 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. Much of the public debate, however, does not revolve around these broader, widely shared topics. Instead, public discourse often focuses on more contentious issues such as evolution, enslavement, racism, and climate change.

We first examine the prevalence of these topics as identified by a large language model, both overall and over time (Figures 5a and A.6). These figures also show that California and Texas textbooks are much more likely to address topics related to climate change than textbooks in our religious school collection. By contrast, our religious school 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 how *often* certain topics are mentioned 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 V.B.5, we can examine differences in the context or sub-topics surrounding a given topic across collections.

First, we examine semantic differences between these topics across collections using topic embeddings regressions, similar to how we did so in Figure 3c. This exercise reveals significantly higher similarities between topics discussed in Texas textbooks and topics discussed in California textbooks than between these public school textbooks and those from our collection of religious school textbooks (with the exception of discussion related to Christianity), indicating that public school textbooks across these states discuss topics more similarly to each other than they do compared to religious school textbooks.

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. We then cluster the word embeddings into subtopics.[11] These figures confirm that many of the topics of interest are discussed in similar contexts across public-school collections which means the collection-specific topic embeddings are close to each other in embedding space, have similar top words, and are a similar distance to each of their top words.

We do see, however, differences in the discussion of *Evolution* across these three collections. As shown in the topic embeddings figures, 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 5a). 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,

[11] We show further validation and topic embedding analysis in Appendices C and E.

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: This is the story of how all was still, all quiet, in silence; there was no movement, no sound, and the whole of the sky was empty. This is the first story, the first speech..."

Exclusive Topics 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, an Indigenous people from what is now Southern California. Only Texas includes topics related to the practice and knowledge of Texas Essential Knowledge and Skills (TEKS) state standards. Only the religious school textbooks have topics related to God, evangelism, and creationism, consistent with what we see in Figure 5a.

Geography and History Another set of values that schools often strive to balance is the extent to which they teach students about their own country's geography and history, versus encouraging awareness and development of their roles as global citizens. Indeed, much of the debate surrounding what children learn involves the knowledge they are given about history and the world around them. As we saw in Figure A.5, geography and history were among the top Social Studies topics identified in each collection. When examining specific mentions of places (Figure 6a), 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 6b shows the most commonly mentioned historical events in our sample of textbooks. A majority were events that occurred in the United States, and the others were wars that the United States was involved with: World War I and World War II. We see varying patterns across the collections, but in particular, we see an increase in mentions of the Civil War over time in the California and Religious collections, but a sharp decrease over the last two decades in Texas. By contrast, we see that Texas and California textbooks are more likely than the Religious collection to mention the American Revolution.

Religious Topics Religion is one of the most frequently cited reasons why some parents opt for private schooling, underscoring the weight placed on its value. Figure A.7 shows that religious topics are present on over 12% of pages in the Religious collection on average as opposed to 4.4% and 2.8% in California and Texas, respectively, and when a specific religion is mentioned, it is overwhelmingly Christianity.

For some topics, however, differences emerge between collections. In Appendix Figure E.6, for example, we see that *Religious* topics are discussed in different contexts across the public school and religious school textbooks. In the Religious collection, the *Religious* topic embedding is closer to word embeddings related to god and Christianity such as savior, pray, creator, sins, Jesus, and heaven. In both the California and Texas collections, the *Religious* topic embedding is closer to word embeddings related to other religions and the U.S. founding era such as Quaker, Islam, Amish, liberties, tolerance, colony, and settlers.

VI.D People

The people who are highlighted in the stories and history presented to children reflect societal values of whose stories should be highlighted and told and how these people are portrayed, and spaces they are expected to hold in society. In this section, we discuss the representation of historic figures and the race, skin color, sexuality, and gender of people who are portrayed to children.

Historic Figures and Occupations. Political and media discussions of curricula are commonly centered around which historical figures and events should be highlighted and how they should be portrayed. 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).

Moreover, examining the representation of historical figures and prominent individuals may have additional benefits, as these real-life examples may serve as tangible role models for students. By highlighting noteworthy figures within educational contexts, schools can offer students aspirational models, potentially influencing their ambitions and shaping their perceptions of what is possible.

We explore the inclusion of historical figures across the collections and over time in Figure A.9 and compare how often they are mentioned across educational settings.[12] The most frequently mentioned famous people in each collection are shown in Table B.2. George Washington, Thomas Jefferson, and Benjamin Franklin – White males who are all Founding Fathers – are among the most commonly included famous individuals across all collections. 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).

Textbooks in our public school collection are slightly more likely to discuss civic leaders (e.g. Martin Luther King Junior, Susan B. Anthony, Frederick Douglass, Rosa Parks) than textbooks in our religious school collection, and this is increasingly true over time. By contrast, textbooks in our religious school collection are more likely to discuss politicians and military personnel (e.g. Abraham Lincoln, John F. Kennedy, Robert E. Lee). In recent years, scientists, explorers, artists, founding fathers, athletes and religious figures are all discussed a similar amount across collections.

Race. A salient part of curricula are the famous individuals who are highlighted for children. 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. To measure how often famous individuals of different races/ethnicities are discussed, we apply Named Entity Recognition (NER) tools as discussed in Section V. One caveat of focusing on famous individuals only to measure the representation of race is that we are unable to capture the race of non-famous individuals in the text. However, we can

[12]We show these patterns by subject for a range of occupations in Figure A.9a. In Social Studies, politicians are the most commonly mentioned, followed by Founding Fathers. In Science, scientists are the most frequently mentioned famous people, but only half of the mentions of famous people are scientists. 4% of the famous individuals in religious school Science textbooks are religious figures. Science books across all collections are more likely to mention artists or athletes than Social Studies books are likely to mention scientists. Public school reading textbooks are more likely to include mentions of artists and athletes, followed by politicians.

look at how often both famous and non-famous individuals of different races are pictured in textbooks by examining the race classifications of detected faces in our sample.

We examine the incidence of famous individuals across the corpora and find that the vast majority of famous individuals mentioned are White individuals (between 81%-89% across all collections and subjects as shown in Table 1). This pattern is consistent over time for both famous White individuals and pictured characters classified as being White (Figure 7). Over time, the Religious collection was more likely to feature famous individuals and pictured characters who are White and less likely to feature famous individuals and pictured characters who are Black relative to the public school collections. Each of the collections has less inclusion of Asian, Latine, or Indigenous famous or pictured individuals, especially relative to the proportion of the Latine student populations in California and Texas. Within racial groups, males are often more likely to be included than females (Figure A.10).

We find that in all collections, most faces depicted are White individuals. However, the difference between the proportions of White faces and Black faces is much closer to parity in faces than it is when examining the race of famous individuals. In public school textbooks, California textbooks have the largest percentage of Asian and Latine pictured characters, though this number is still relatively small; 10-12% each compared to 7-10% each in Texas public school textbooks and 5-6% each in the Religious collection (Table 1 and Figure A.10b).

Skin Color. Skin color is one of the first ways that children learn to distinguish humans, and is distinct from measures of putative race but is highly related to racial classification (Kessen and Weiskopf, 1976; Katz and Kofkin, 1997; Kinzler and Spelke, 2011; Adukia et al., 2025). In Figure 7c, 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 7d 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.

Birthplace. When we examine the birthplaces of the famous individuals to whom children are exposed (Figures 6c 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.

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.4, 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. While the sentiment of this passage is positive overall, by not discussing any LGBTQIA+ individuals of color or who identify as female, this passage fails to take an intersectional lens. 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 over time, we see the public school collections have higher female representation in both the text and the images of textbooks than the Religious collection (Appendix Figure A.12). 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. Notably, however, in public school Reading textbooks, females are mentioned as frequently 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.

Gendered Context. We then measure the contexts in which females and males are portrayed in both the text and images in 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). We describe this approach in Section V.B.5.

Table 2 shows the top 10 relative words most closely associated with females vs. males and, separately, with males vs. females for 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 positive on average, 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 8a and 8b.

In Figures 8c and 8d 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, and household) is more skewed towards females and the *Politics* domain (which includes words such as government, democracy, and political) 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. We find a similar (though smaller) gender skew in the *Politics* domain towards males for the Texas and Religious textbook 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.

VII 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: first, public-school textbooks adopted in California and Texas are often similar to each other on a variety of dimensions, and second, textbooks religious publishers' textbooks 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.

VII.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.

VII.B Conceptual Framework

The framework emphasizes three market and regulatory forces which influence the content of textbooks across states.

Principle of Minimum Differentiation 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).[13] In our setting, interpret "position" as a one-dimensional index capturing overall textbook positioning (or narrative emphasis) along a salient dimension of content and purchasers are local school districts. When prices are fixed, regulated, or otherwise second-order in procurement, this benchmark predicts that competing publishers aiming to serve the national public-school market will tend to offer very similar textbooks that appeal to the median purchaser (e.g. school district).

State Standards and Costly Customization as Constraints. To capture how state standards can impact textbook publication, 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". Even when California and Texas are treated symmetrically (equal-sized markets), costly

[13] In the classic two-stage Hotelling model with endogenous prices, the minimum-differentiation conclusion is not generally robust. In particular, the price subgame may fail to admit a pure-strategy equilibrium when firms locate too close, and modifications can instead deliver forces toward greater differentiation. This is why we use a pure content-positioning benchmark as a reduced-form representation of adoption environments where price competition is muted. See, e.g., Eaton and Lipsey (1975); d'Aspremont et al. (1979); Economides (1984); Osborne and Pitchik (1987).

customization can therefore generate cross-state spillovers in what is covered.

Product Differentiation for Different Markets 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., Mussa and Rosen, 1978; 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.

VII.C Empirical Evidence

The predictions generated by these market and regulatory forces are consistent with our main findings in Section VI. Namely (1) Texas and California textbooks are more similar to each other than we might have expected given differences the narrative of political polarization and (2) textbooks from Religious collection differ in ways that are consistent with differences in beliefs between religious and non-religious individuals in the United States. Below, we provide additional examples of content spillover and product differentiation across states in response to changes in state content standards consistent with these predictions. These exercises are descriptive and are intended to evaluate whether the patterns in the data line up with the framework's directional predictions.

VII.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. We focus on three cases where one state adopts a new set of standards in a given subject (Texas reading, California science, and Texas social studies). For each case, and separately by grade (3 and 5), we construct a transparent "new-standards" dictionary from the standards text: the set of tokens that appear in the adopting state's updated standards but not in its prior standards, excluding tokens that already appear in the other state's pre-period standards. We then measure, for each textbook, the share of words that belong to this dictionary and track how this share evolves by year-in-use in California and Texas (Figure 9). The solid vertical line marks the standards-change year and the dashed vertical lines mark each state's subsequent

textbook adoption cycle in that subject.

The figures line up with the spillover logic in Equation (1). In the state that revises its standards, the share of new-standards terms rises sharply after the next adoption cycle (a direct effect). In several cases, the non-adopting state also exhibits an increase after its own subsequent adoption cycle (a spillover), even though its standards did not change. To summarize these descriptive patterns, we estimate the following book-level regression:

$$(2) \quad y_{bcsg} = \beta^D \, \mathbf{1}\{Post_{bcsg} = \text{Direct Effect}\} + \beta^S \, \mathbf{1}\{Post_{bcsg} = \text{Spillover}\} + \alpha_{c\times s\times g} + \varepsilon_{bcsg}$$

where y_{bcsg} is the share of words in textbook b from collection c, subject s, and grade g that fall in the relevant new-standards dictionary. The omitted category is the share of words before the standards change and a new textbook is adopted."Direct Effect" denotes post-adoption textbooks in the state that revised its standards, and "Spillover" denotes post-adoption textbooks in the other state. We include collection×grade fixed effects in subject-specific regressions (and collection×grade×subject fixed effects in the pooled specification) and cluster standard errors by book. Table 3 shows large and precisely estimated direct effects and smaller, but generally positive, spillovers for Reading and Social Studies textbooks.

Taken together, these exercises provide empirical evidence that changes in state standards translate 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 VI.

VII.C.2 Product Differentiation by Religious Publishers

The conceptual framework in Section VII emphasizes that religious publishers are not competing for adoption in the same national public-school market. Instead, they serve private-school and homeschool consumers who face different constraints (no state adoption screens) and plausibly have different preferences. The segmented-market logic therefore predicts systematic differentiation of religious textbooks along dimensions that are salient to these consumers, rather than convergence toward a common "median district" product. In this section we provide descriptive evidence that several of the key ways in which Religious-collection textbooks differ from California and Texas textbooks line up with differences in the demographics and beliefs of the populations these books are designed to serve.

Demographics and racial representation. Religious-collection textbooks contain a higher share of pictured characters and references to famous individuals who are White, and detected faces are lighter on average (Table B.2). These patterns partly reflect differences in student demographics across school types. Figure 10 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, imagery, and historical/civic reference points toward the modal student audience in its target market.

Beliefs, values, and "character education." If religious publishers are differentiating to match the values of their consumers, 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,[14] (ii) non-practicing respondents in California, and (iii) non-practicing respondents in Texas.

First, Figure 12a uses a long-running GSS module (1983–2018) asking respondents to choose the most important trait to teach children. 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. Figure 12a shows that 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.4. This figure also shows that while most character values are more likely to be found in our collection of Religious textbooks, values such as creativity and understanding are the exception; those are more likely to be found in California and Texas textbooks.

[14]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.

Religion, science beliefs, and gender norms. The same pattern appears for other salient beliefs. Figure 11 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 and prayer is among the most prominent topics in reading materials. Figure 11 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 and more traditional gender associations in text and images (see Table B.3 and the gender-skew and sentiment patterns discussed in Section VI.D).

Taken together, these descriptive exercises support the segmented-market interpretation: 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.

VIII Conclusion

What we teach our children matters. 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. By 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. Over one-third of each of the textbook collections communicate character values, with religious school textbooks doing so at higher rates. Notable similarities also appear: 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, and present the U.S. founding era and slavery in comparable ways.

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.

An important consideration for the implications of our work is that curricula can take differing forms: curricula as written, curricula as implemented (by teachers), and curricula as discussed (among students). In our setting, we focus our analysis on the first of these, which is curricula as written. 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. This suggests that the written curriculum matters and the importance of analyzing textbook book content remains despite potential difference 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.

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.

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.

Morris, Richard. 1973. Seven Who Shaped our Destiny: The Founding Fathers as Revolutionaries. Harper & Row.

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.

Mussa, Michael, and Sherwin Rosen. 1978. "Monopoly and Product Quality." Journal of Economic Theory, 18(2): 301–317.

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.

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.

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.

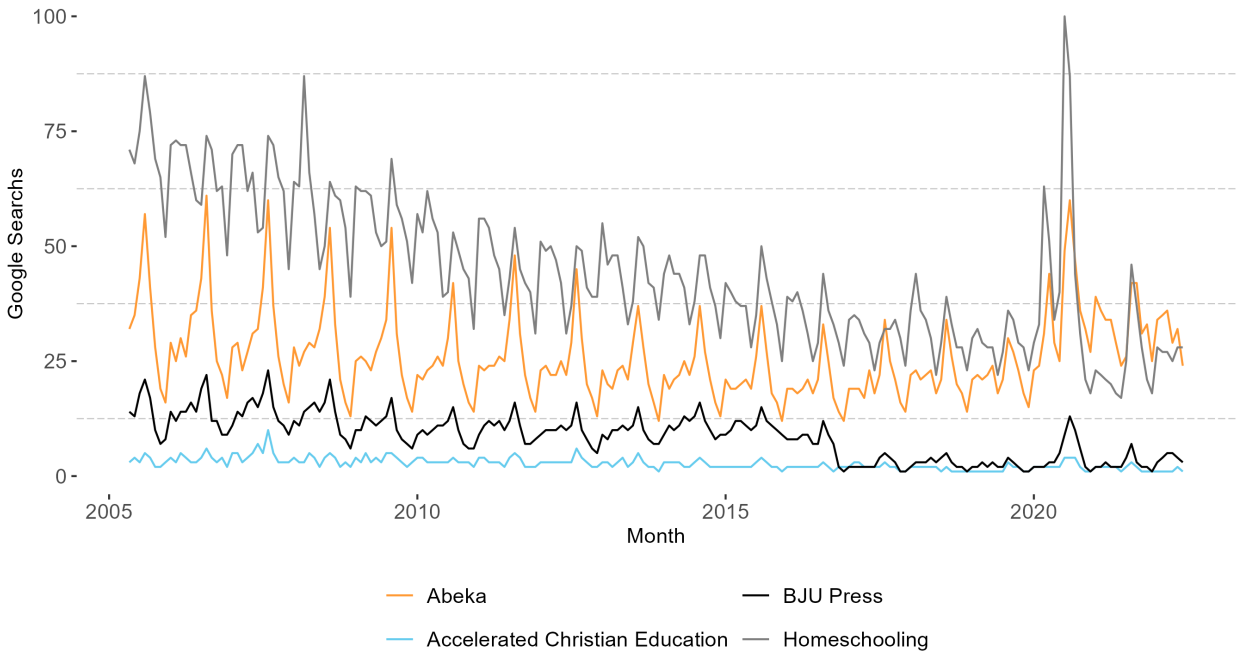
Wheaton, Brian. 2020. "Laws, beliefs, and backlash." Working Paper.

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.

IX Main Figures

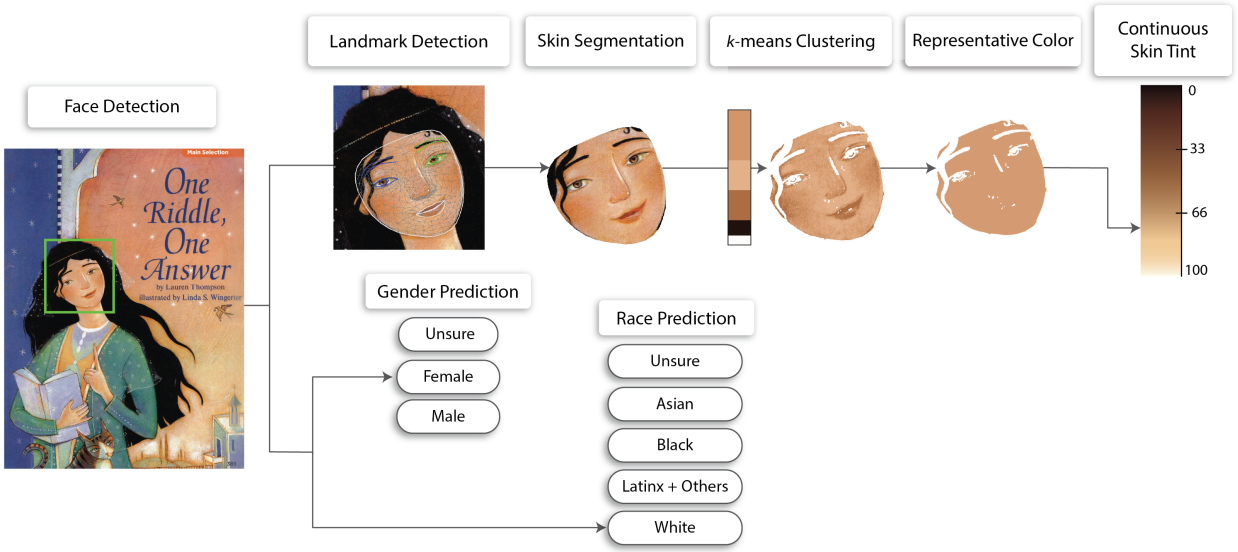
Figure 1: Google Search Interest Over Time — Homeschooling and Religious Collection



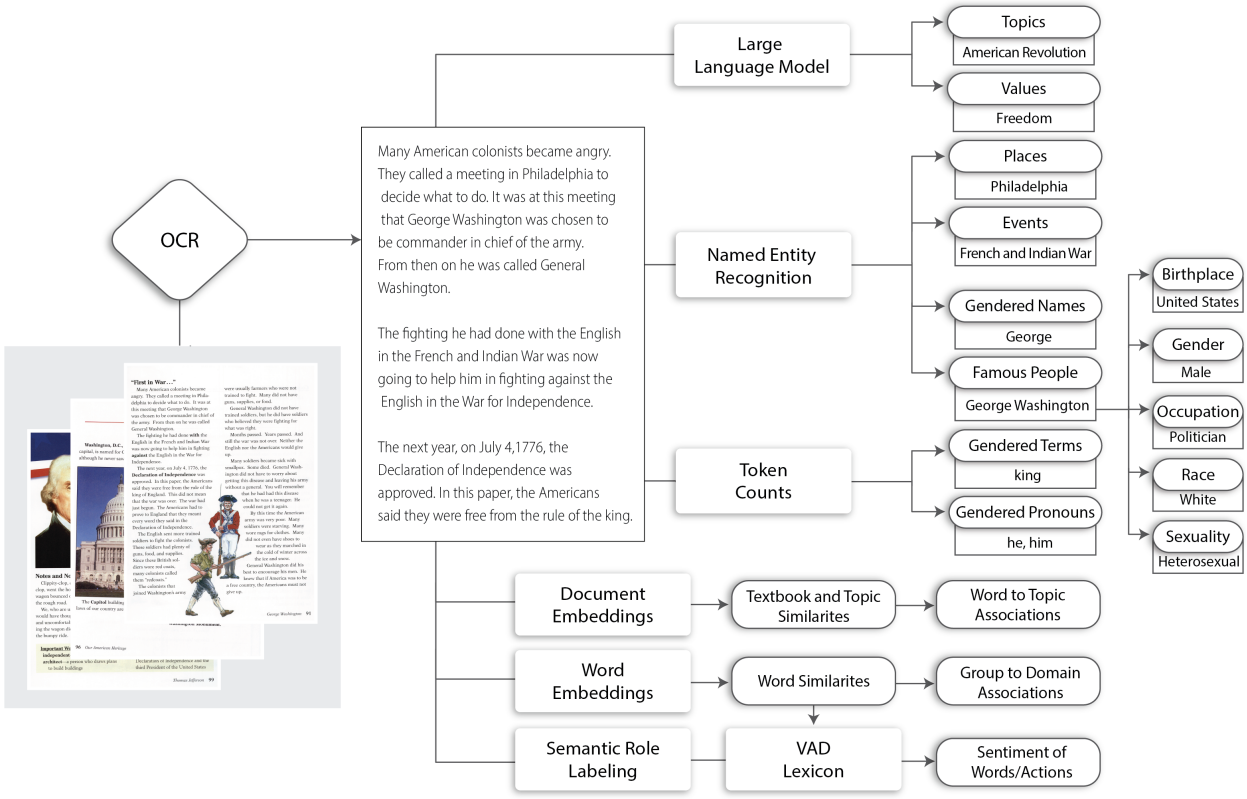
Notes: This figure illustrates Google search interest from Google Trends data for the search term homeschool and each of the education publishing companies in the religious school textbook collections; Abeka, Acceleration Christian Education (ACE), and BJU Press (BJU). Note that we use the company topics 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 2: 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.

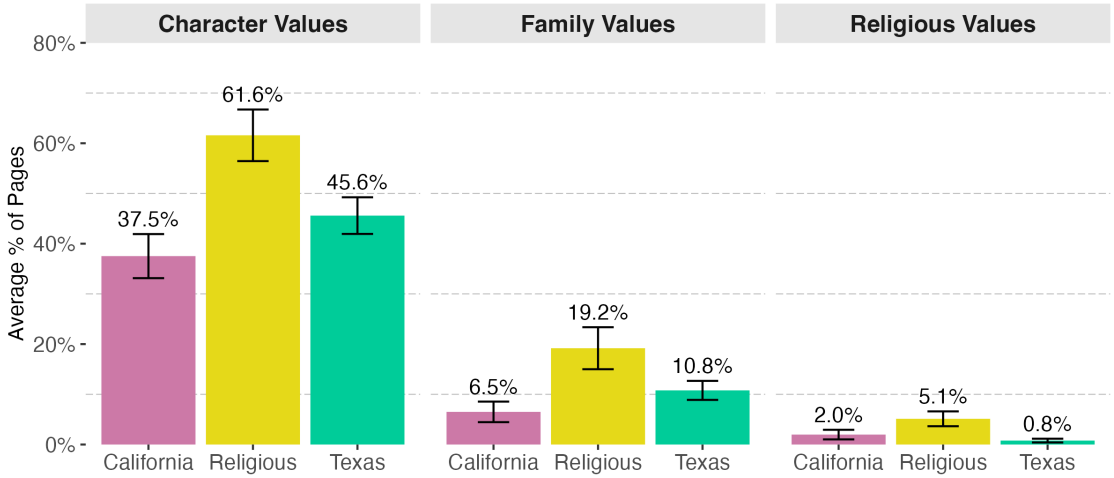
Figure 3: Textbook Embeddings



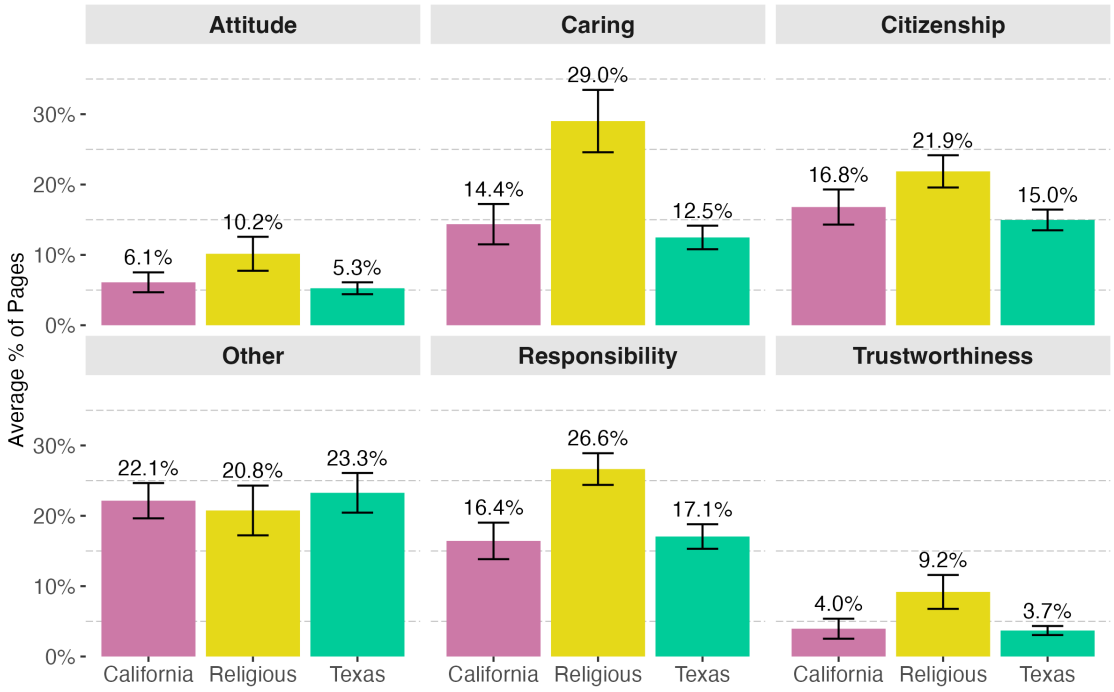
Notes: In panel (a) and (c), we show analyses using textbook-level embeddings estimated using the text from each textbook. In panels (b) and (d), we show textbook-level embeddings constructed by averaging across image embeddings in a textbook. Each image embedding is estimated from an AI generated description of a detected character's setting, location, and actions as described in Section V.B.5. 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 4: Values

(a) Character, Family, and Religious Values by Collection



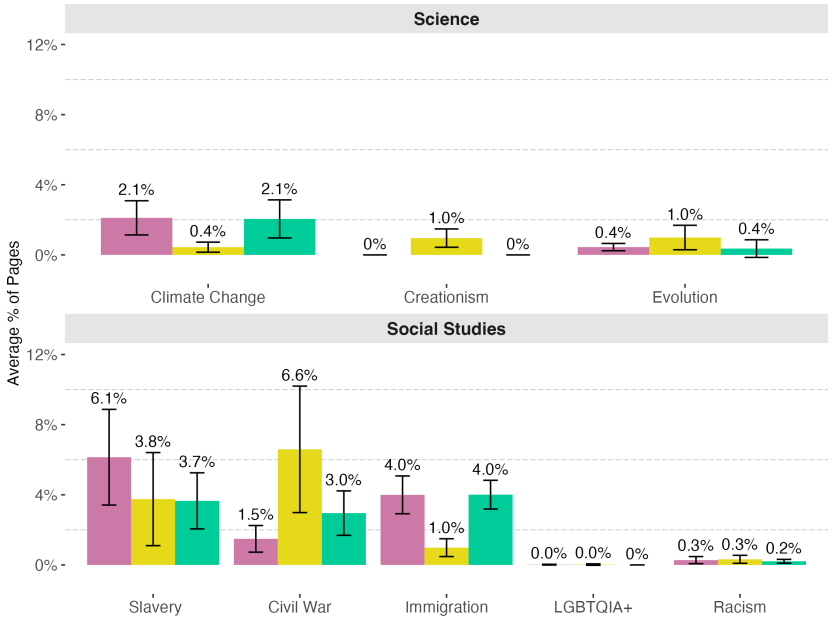
(b) Average Percent of Reading Pages Containing Character Values Categories



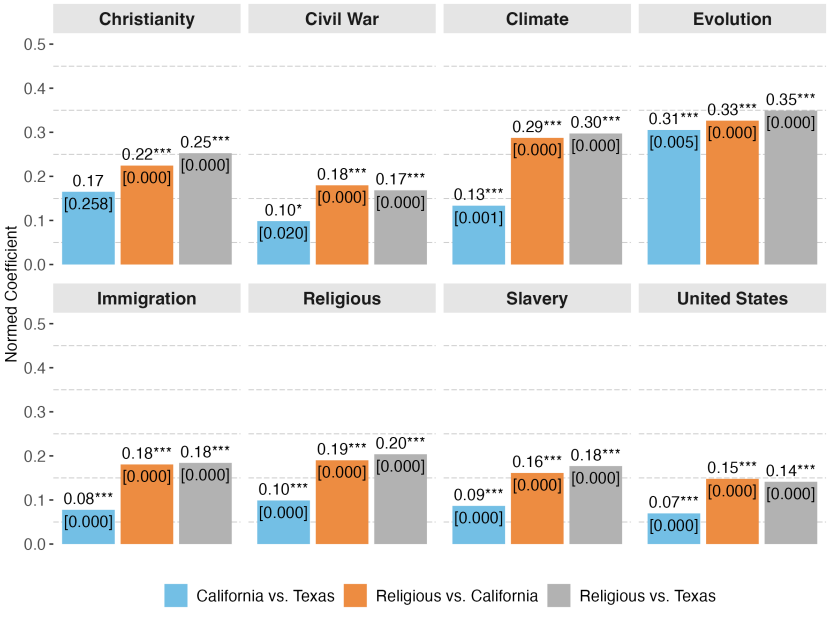
Notes: Panel (a) reports the average share of pages that contain character values by collection and subject. Each collection *times* subject average is weighted where the weighting unit is defined as a textbook × grade × year-in-use. Standard errors are clustered at the textbook-level. Error bars show 95% confidence intervals. Panel (b) disaggregates the broader category of character values into specific subcategories outlined in state standards as described in Appendix Table B.1.

Figure 5: High-Profile Topics

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



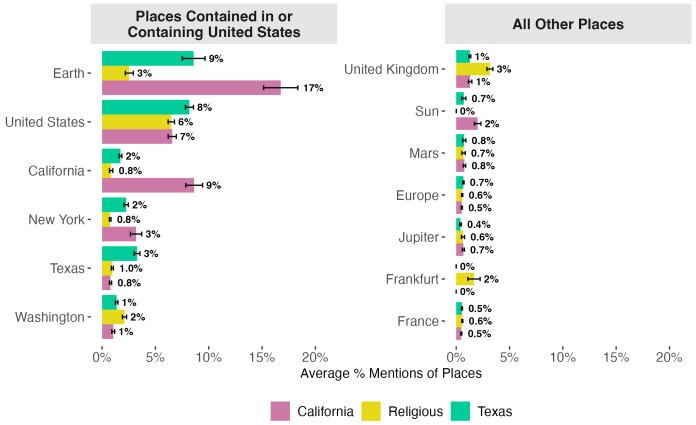
(b) Topic Embedding Regressions



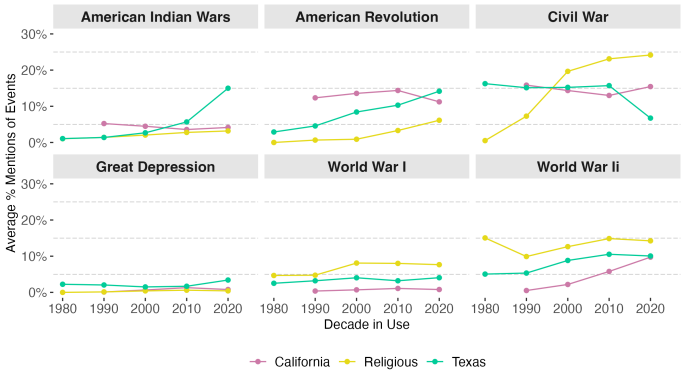
Notes: Panel (a) reports average share of pages containing a specified topic in a given collection and subject. Each collection *times* subject average is weighted where the weighting unit is defined as a textbook × grade × year-in-use. Standard errors are clustered at the textbook-level. Error bars show 95% confidence intervals. Topics are identified on pages using an LLM. 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 6: Geography and History

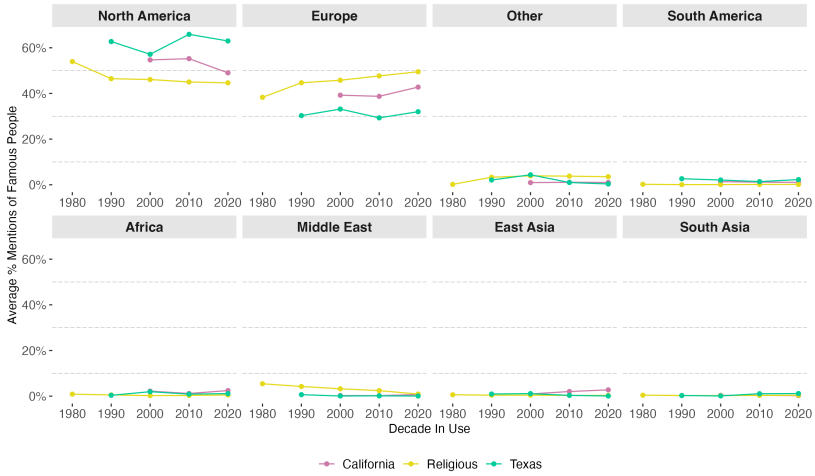
(a) Geography



(b) Historical Events



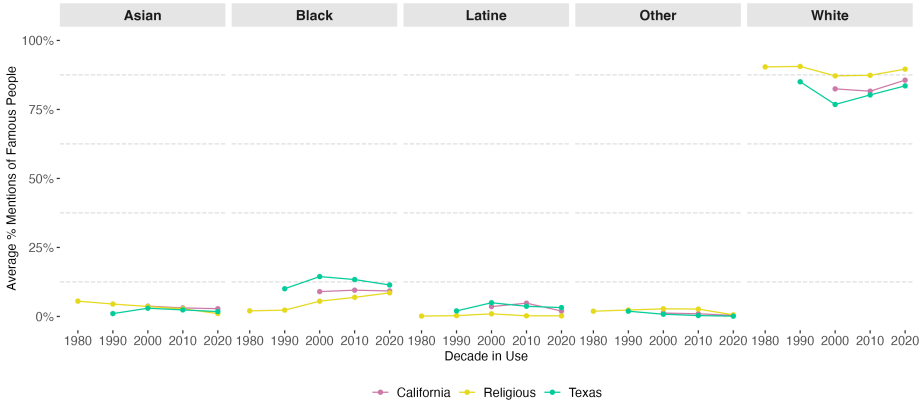
(c) Birthplace of Famous Individuals



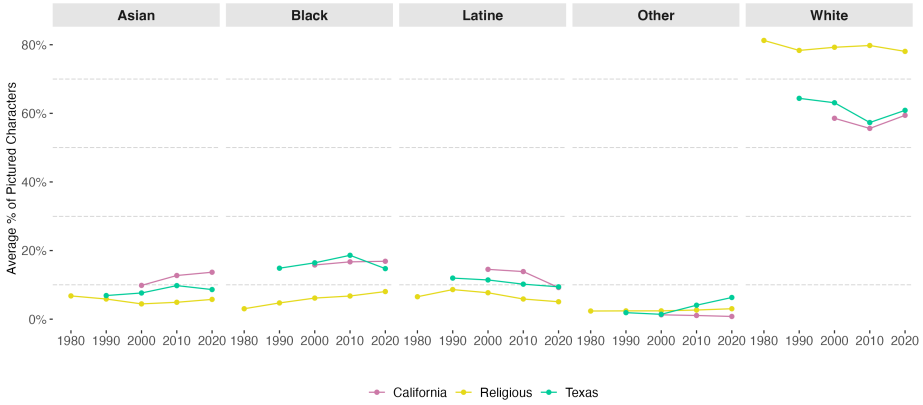
Notes: Panel (a) illustrates the most frequently mentioned locations by collection and decade-in-use. Each collection *times* decade-in-use average is weighted where the weighting unit is defined as a textbook × subject × grade × year-in-use. Standard errors are clustered at the textbook-level. Error bars show 95% confidence intervals. Similarly, Panel (b) shows the most frequently mentioned events by collection over time and Panel (c) shows the mentions of famous people by their country of birth. We determine mentions of locations, events, and famous people using Named Entity Recognition.

Figure 7: Representation of Race and Skin Color in Textbooks

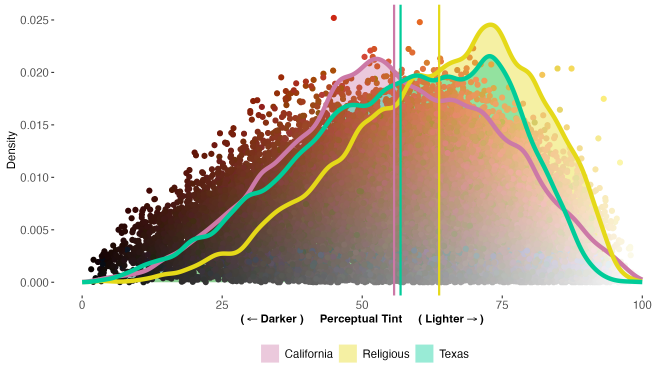
(a) Race (Text)



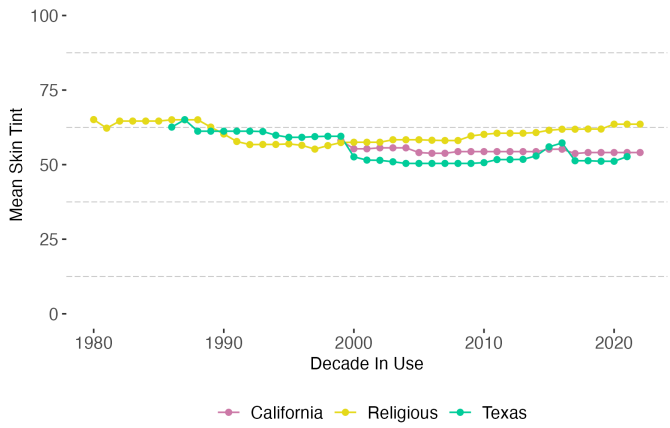
(b) Race (Images)



(c) Skin Tint Distributions



(d) Skin Tint Over Time

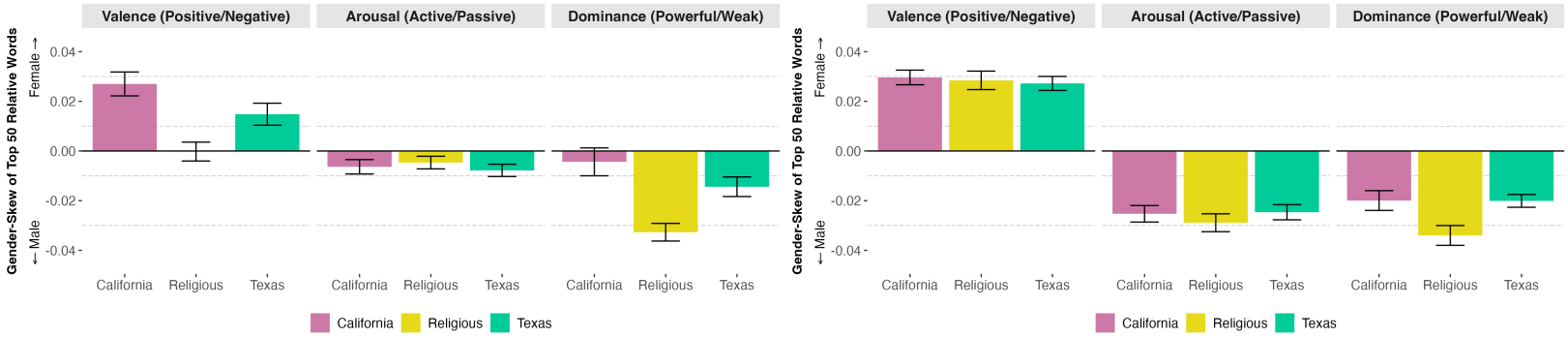


Notes: This figure shows the presence of predicted putative race and skin color. Panel (a) shows the average percent of mentions of famous people in the text that are of a specific race in each collection by decade in use. Panel (b) shows the average percent of pictured characters classified as a specific race in each collection by decade in use. Each collection *times* decade-in-use average is weighted where the weighting unit is defined as a textbook × subject × grade × year-in-use. Standard errors are clustered at the textbook-level. Error bars show 95% confidence intervals. Panels (c) and (d) show our analysis of the representative skin colors of the faces we detect in images, focusing on faces considered to be human skin colors (polychromatic skin colors where $R \geq G \geq B$). 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 these patterns over time.

Figure 8: Gender Sentiment (VAD) and Skew by Domains

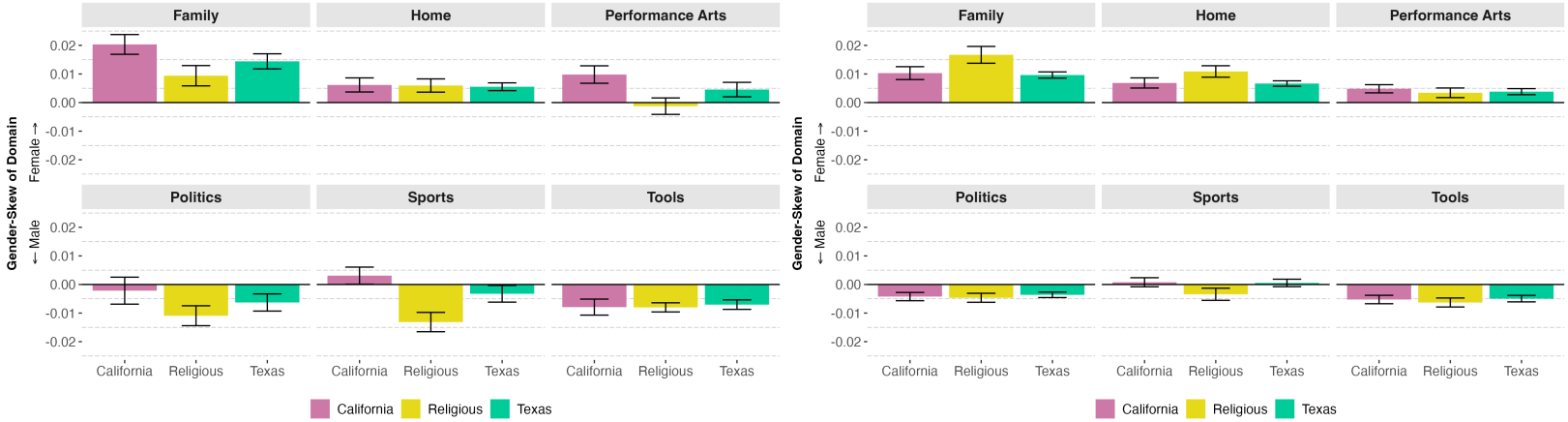
(a) Sentiment via VAD (Text)

(b) Sentiment via VAD (Images)



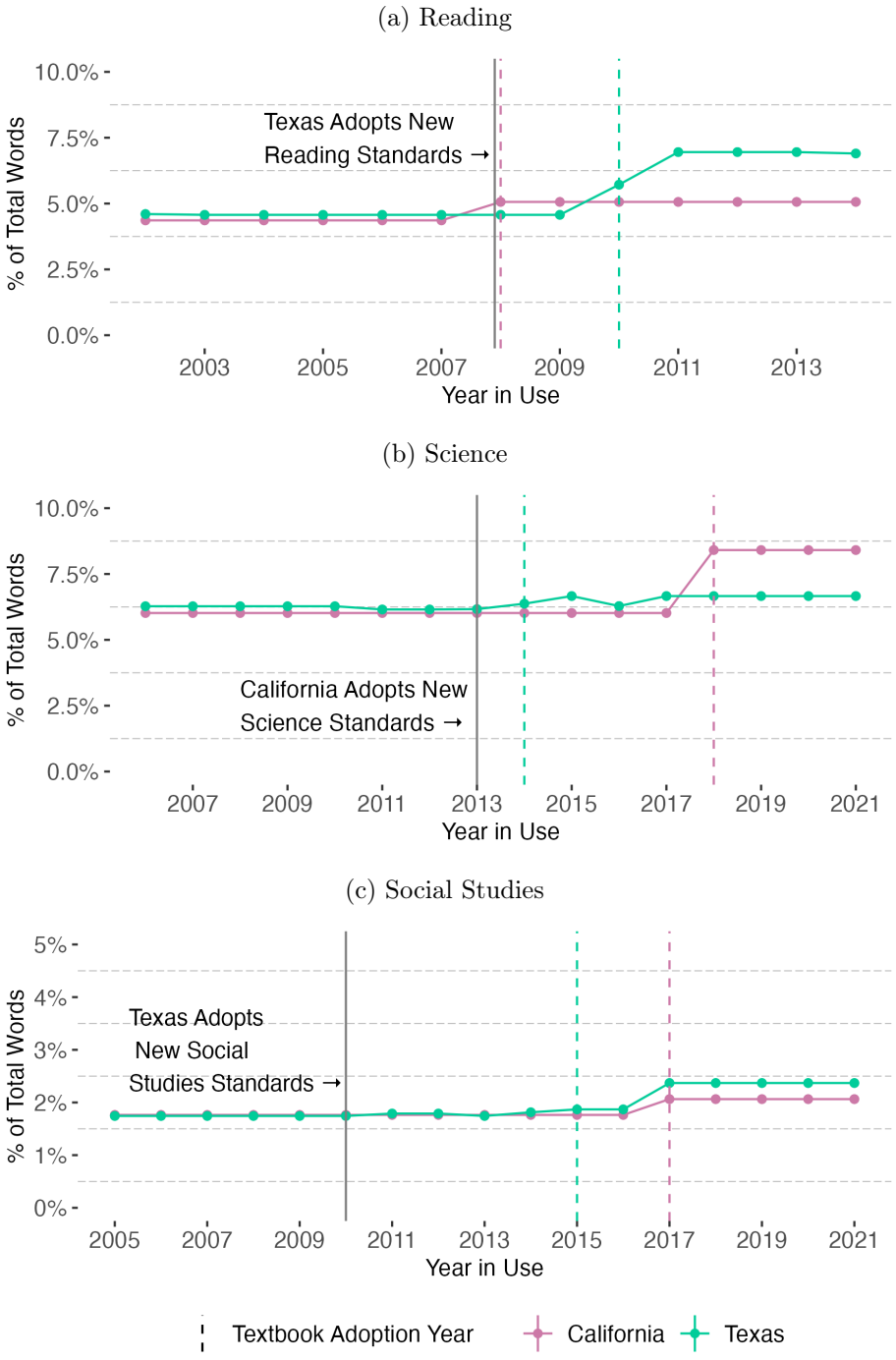
(c) Gender Skew by Domain (Text)

(d) Gender Skew by Domain (Images)



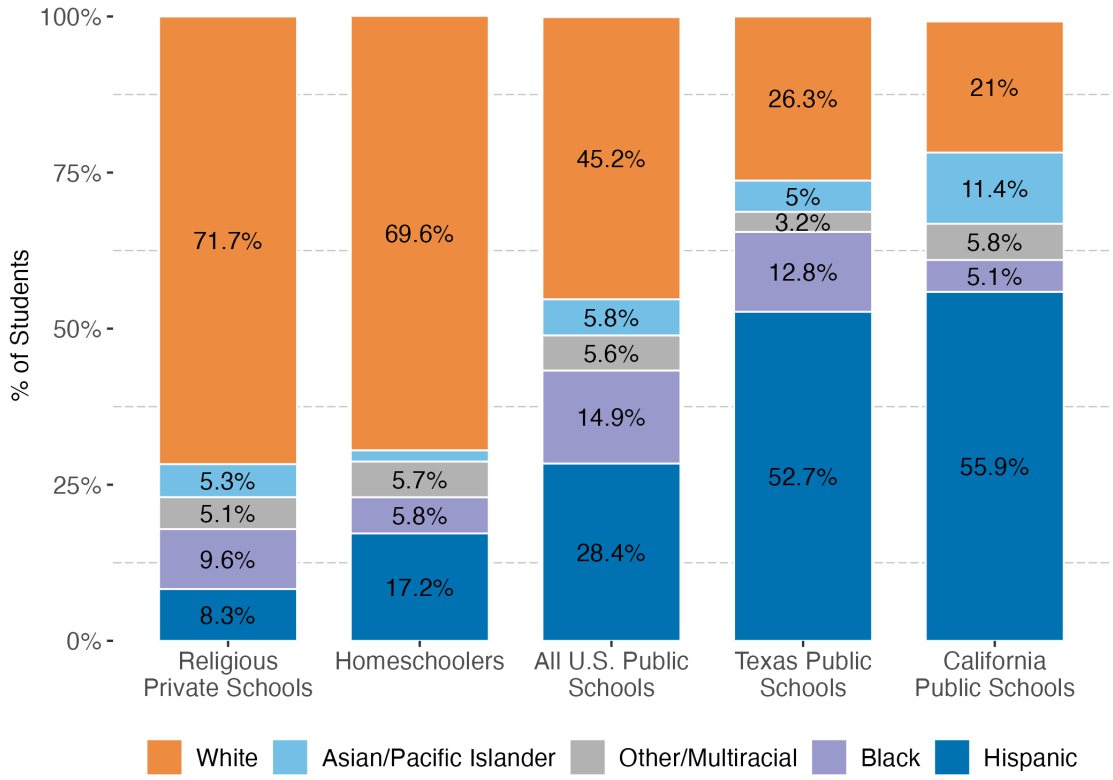
Notes: Bars report collection-specific coefficients estimating differences between males and females from embedding-based regressions; error bars denote 95% confidence intervals, with standard errors clustered at the textbook level. Panels (a) and (b) show gender differences in the average Valence, Arousal, 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 9: Standards-Driven Vocabulary in Textbooks



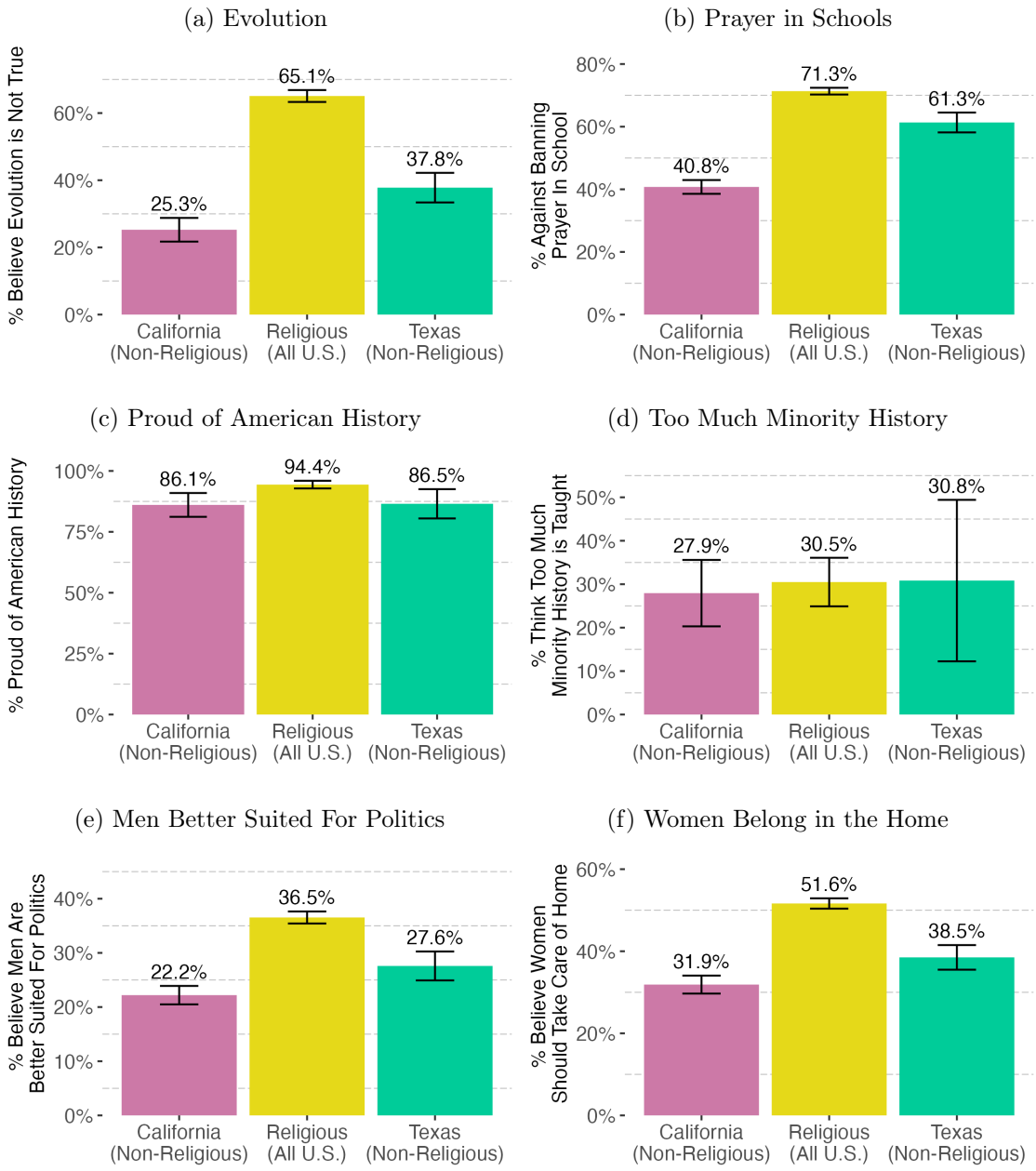
Notes: Each panel plots, by year in use, the share of words in California and Texas textbooks that belong to a dictionary of "new-standards" terms. For each subject and grade, we define this dictionary as new words that appear in the adopting state's revised standards but not in its prior standards and 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 solid vertical line marks the standards-revision year, and the colored dashed lines mark each state's subsequent textbook adoption cycle. 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 10: Racial/Ethnic Composition of K–12 Students by School Type



Notes: Data complied 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.

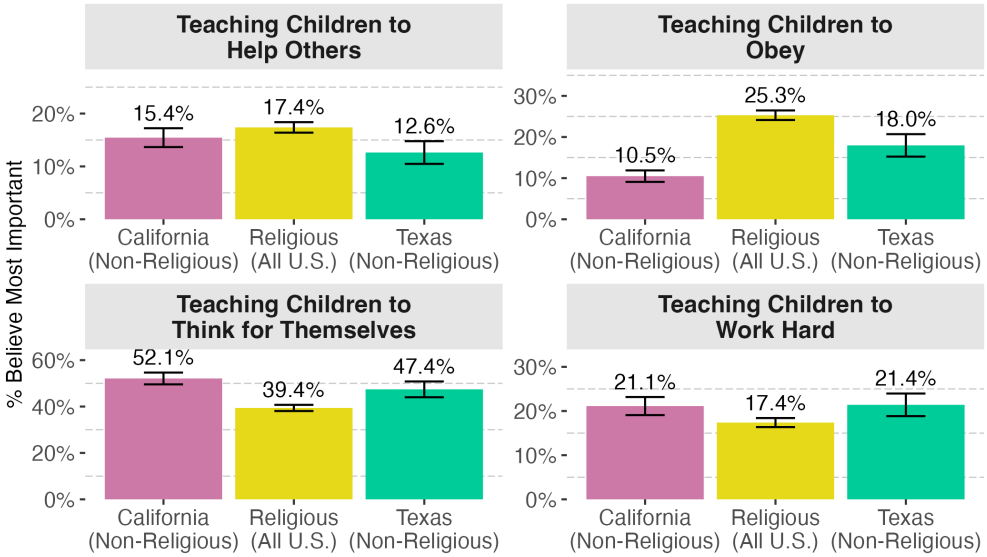
Figure 11: Beliefs by State and Religiosity



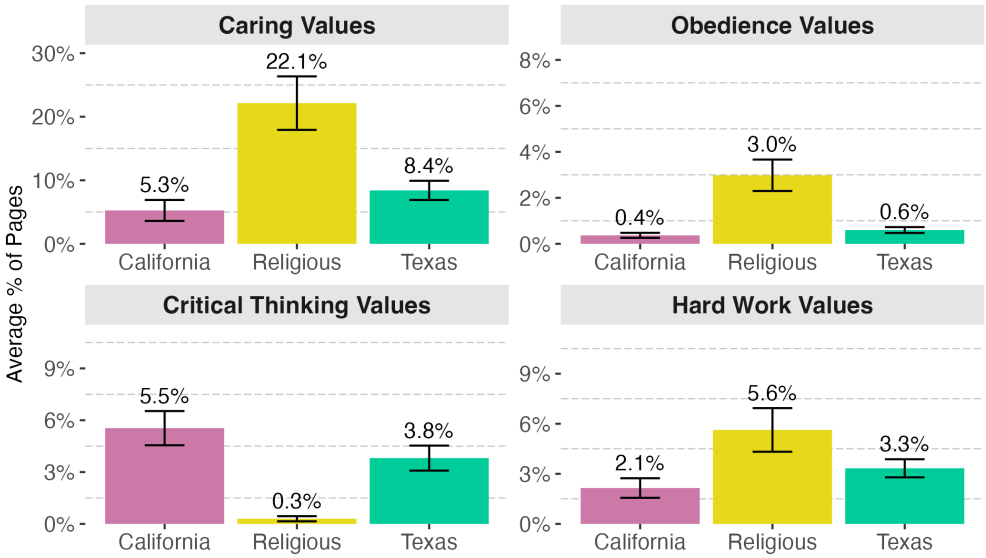
Notes: Bars report survey-weighted means from pooled General Social Survey cross-sections, using all years in which each question was asked; error bars denote 95% confidence intervals that account for the GSS survey design. The "Religious (All U.S.)" group consists of respondents nationwide who attend 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 the available GSS evolution/creationism items.

Figure 12: 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 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. Panel (b) shows the average share of pages that contain specific values related the the GSS question in Panel (a). Each collection average is weighted where the weighting unit is defined as a textbook × subject × grade × year-in-use. Standard errors are clustered at the textbook-level. Error bars show 95% confidence intervals.

X 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 Studies 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	7	7
Average number of values per page	6	6	6
Weighted Average: Skin Color			
Average skin tint of faces	55	62	57
Weighted Average Percent: Putative Race of Characters			
Average percent of pictured characters classified as Asian	10%	5%	7%
Average percent of pictured characters classified as Black	15%	6%	15%
Average percent of pictured characters classified as Latine	12%	6%	10%
Average percent of pictured characters classified as White	52%	70%	57%
Average percent of pictured characters classified as other race	1%	2%	2%
Average percent of pictured characters classified as unsure race	10%	11%	9%
Weighted Average Percent: Putative Race of Famous People			
Average percent of famous people classified as Asian	3%	4%	2%
Average percent of famous people classified as Black	9%	5%	12%
Average percent of famous people classified as Indigenous	1%	2%	1%
Average percent of famous people classified as Latine	4%	0%	3%
Average percent of famous people classified as White	83%	89%	81%
Weighted Average Percent: Gender of Pictured Characters			
Average percent of pictured characters classified as female	34%	30%	36%
Average percent of pictured characters classified as male	56%	66%	57%
Average percent of pictured characters classified as unsure gender	10%	5%	8%
Weighted Average Percent: Gendered Words and Famous People			
Average percent of female gendered words	39%	36%	45%
Average percent of famous people classified as female	16%	17%	20%
Weighted Average Percent: Character Type			
Average percent of pictured characters classified as human	67%	66%	70%

Notes: For all collection averages, we take a weighted average where the weighting unit is defined as a textbook × subject × grade × year-in-use. This methodology ensures that textbooks used over longer periods are given greater weight and that equal emphasis is placed on each subject and grade. Average number of topics and values per page on conditional on having at least one topic/value identified on a page.

Table 2: 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 3: 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	0.0168***	0.0188***	0.0188***	0.0042**
	(0.0021)	(0.0031)	(0.0032)	(0.0017)
Spillover	0.0048***	0.0079***	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	0.392	0.429	0.431	0.356

Notes: This table reports estimates of Equation (2). The dependent variable is the share of words in a textbook-year that belong to a subject–grade "new-standards" dictionary, defined as tokens introduced by the adopting state's revised standards relative to its prior standards and not present in the other state's pre-period standards. "Direct Effect" indicates post-adoption textbook-years in the standards-changing state, and "Spillover" indicates post-adoption textbook-years in the other state. The omitted category is textbook-years prior to the post-revision adoption cycle. Column (1) pools subjects; Columns (2)–(4) are subject-specific. 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

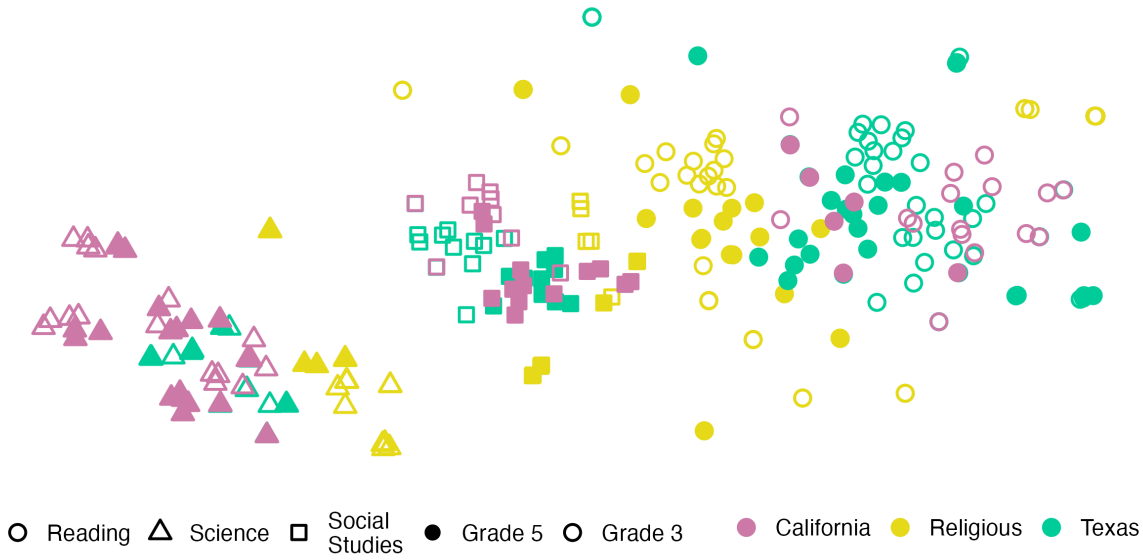
Anjali Adukia, University of Chicago, MiiE Lab, IZA, NBER
Emileigh Harrison, University of Chicago, MiiE Lab, Inclusive Economy Lab

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

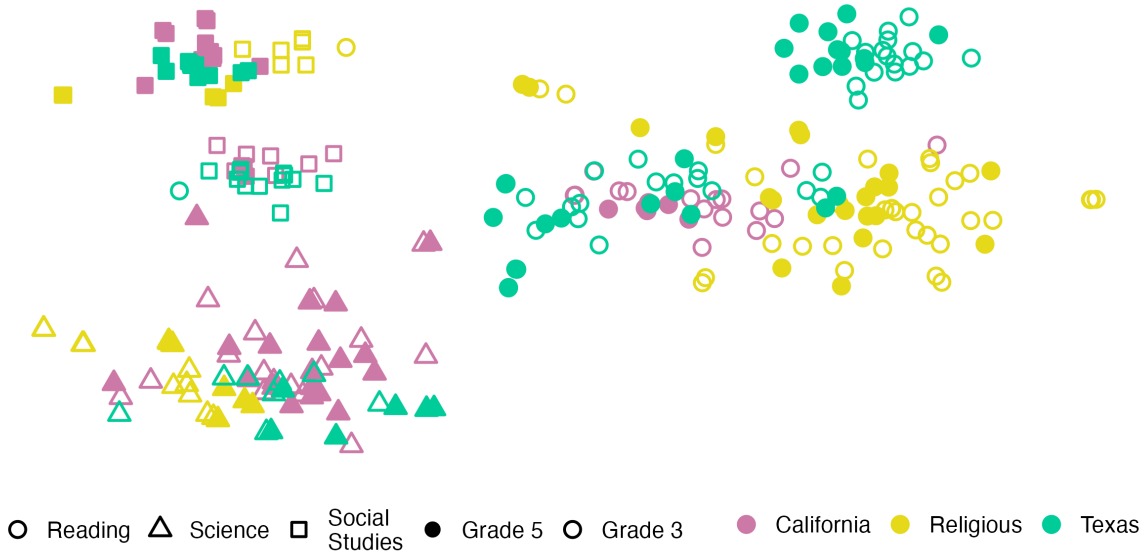
A Appendix Figures

Figure A.1: Textbook Embeddings

(a) Text



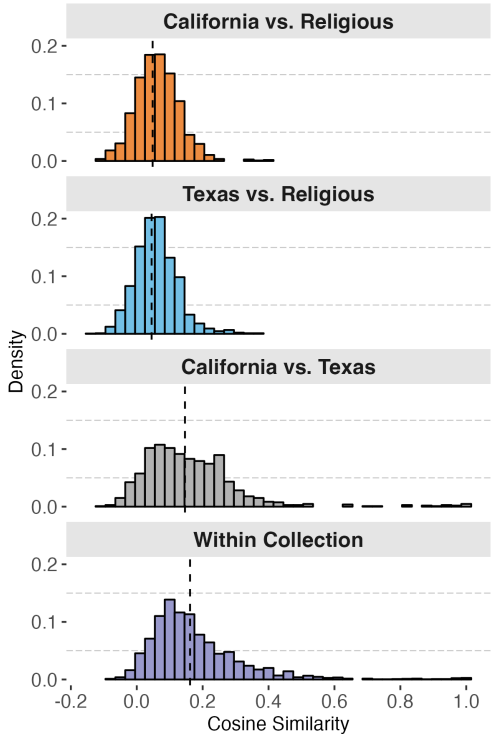
(b) Images



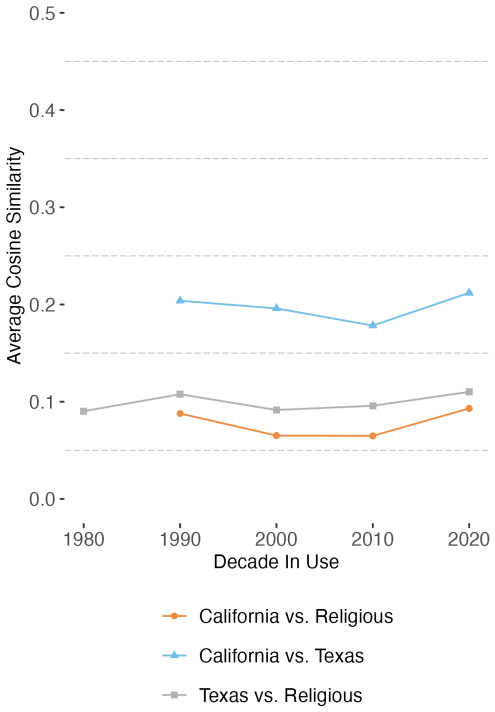
Notes: Each observation represents a book from a specific collection (color), subject (shape), and grade (outline). The filled shapes indicate 5th grade textbooks and hollow shapes indicate 3rd grade textbooks. Their content, originally represented in a 300-dimensional space, is here reduced to 2D via 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.2: Average Cosine Similarity Between Book Embeddings by Collection Pairs

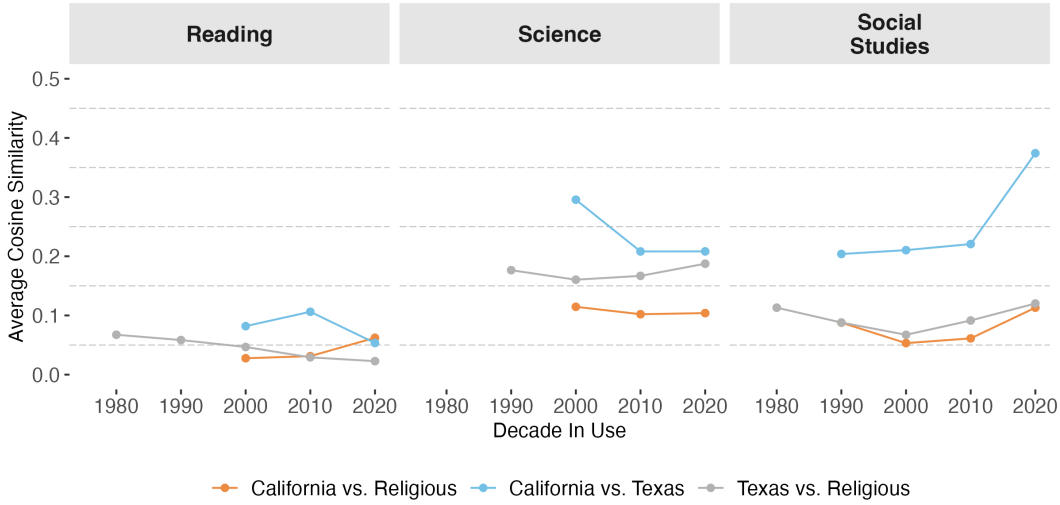
(a) Histogram of Similarities Between Book Embeddings by Collection Pairs



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



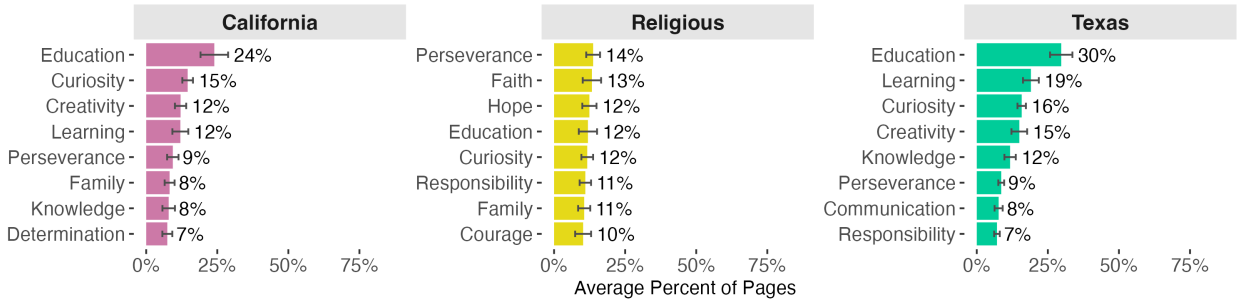
(c) Average Similarity Between Book Embeddings by Collection Pairs and Subject over Time



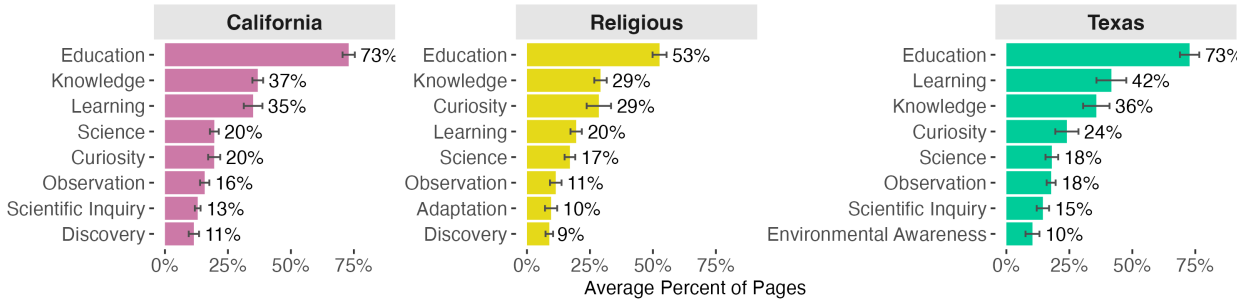
Notes: In Figure (a), we plot a histogram of the cosine similarities between two books that share the same subject, grade, and year in use. We separate the densities by collection pairs (e.g. one California textbook and one Texas textbook) and plot within collection book pair similarities for comparison. 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.3: Top Values by Collection

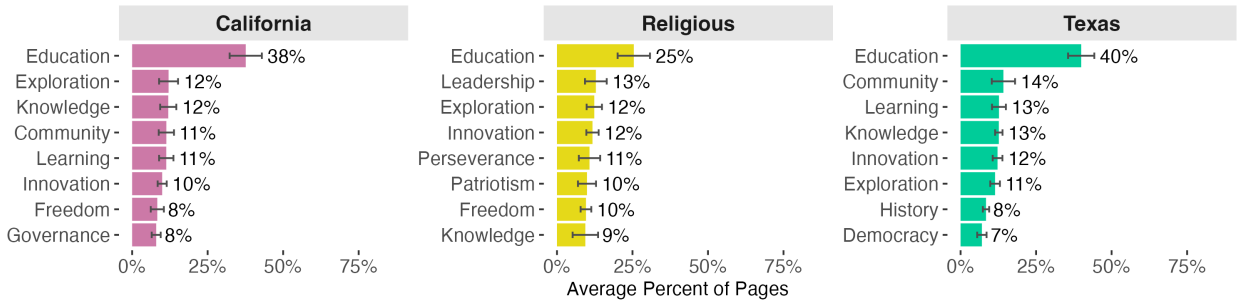
(a) Reading



(b) Science



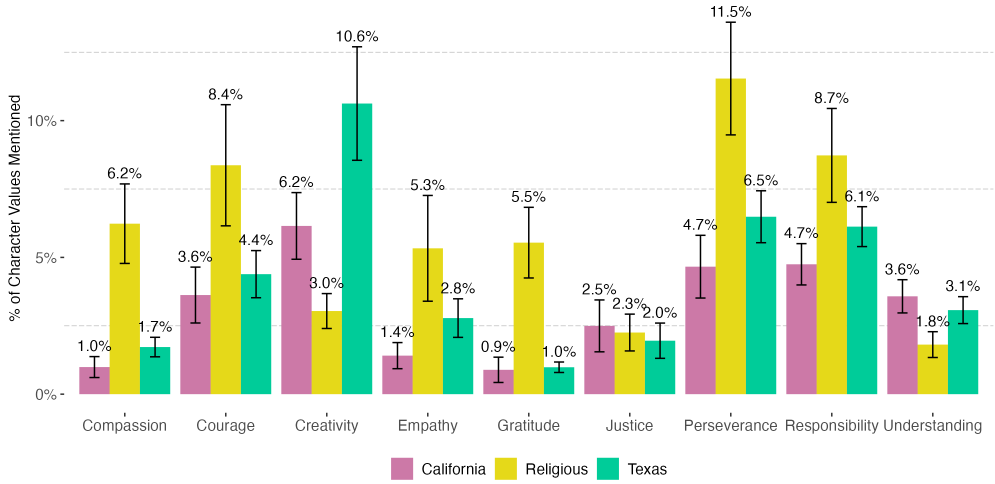
(c) Social Studies



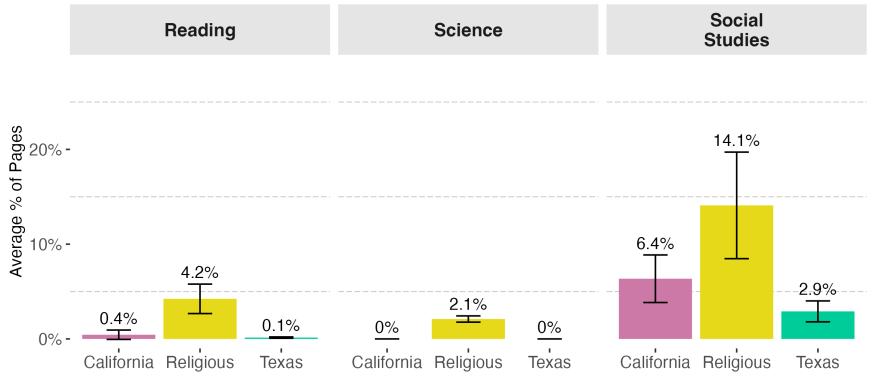
Notes: This figure shows the top values (e.g. values that appear on the highest average percentage of pages) within a collection. To calculate the average percentage of pages containing each value, we first calculate the percentage of pages for each collection, subject, grade, and year in use. We then average these percentages by collection, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages.

Figure A.4: Character Values and Religious Values

(a) Specific Character Values Mentioned by Collection



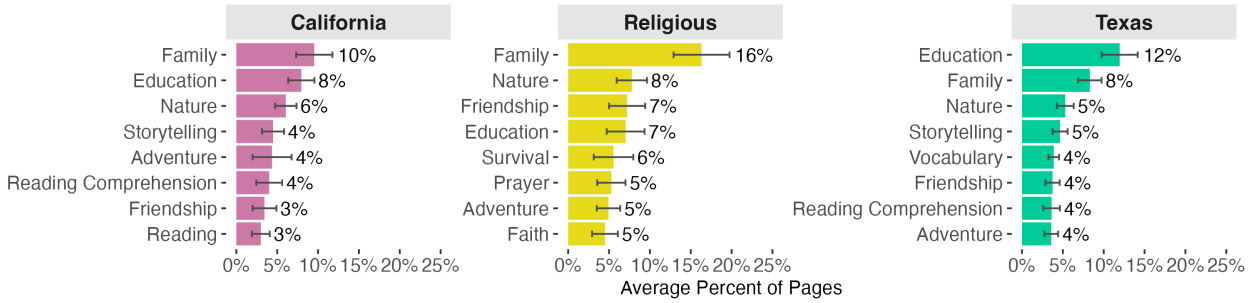
(b) Average Percent of Pages Containing Religious Values by Subject



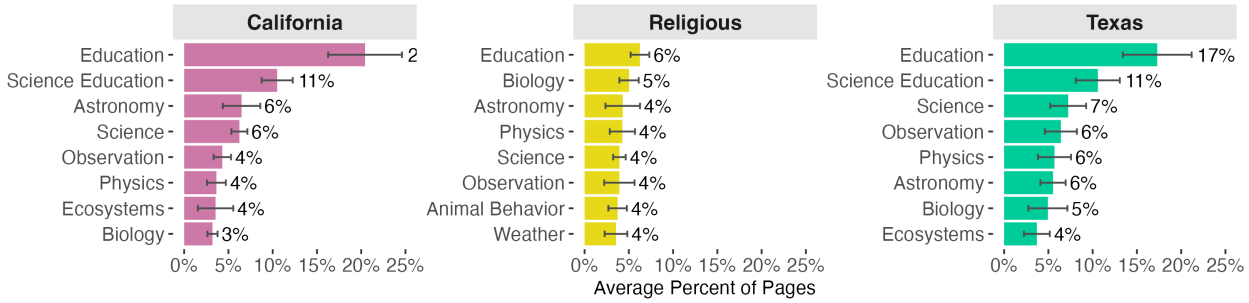
Notes: This figure shows (a) the proportion of all character value mentions that is accounted for by the most frequently mentioned values in each collection by subject and (b) the average percentage of all Reading pages in a collection that contain character, family, moral, or religious values. To determine the average percentage of pages, we first calculate the percentage of pages within each collection, subject, grade, and year in use. We then average these percentages at the collection level to ensure that pages from each grade, subject, and year in use are weighted equally. Values were classified by a large language model.

Figure A.5: Top Topics by Collection

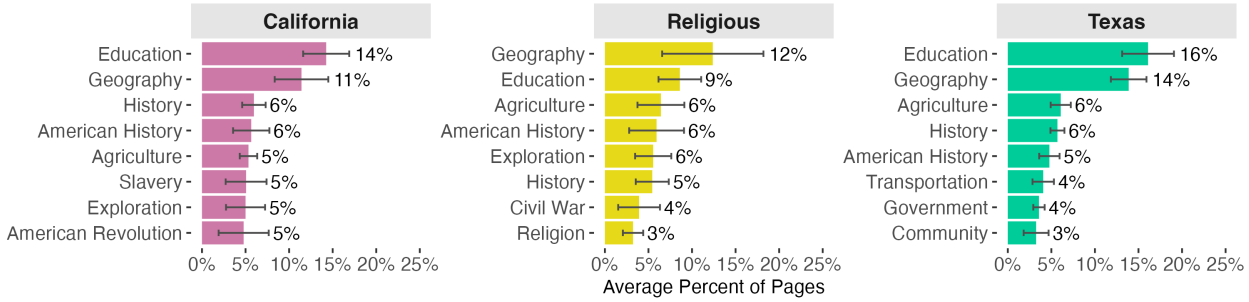
(a) Reading



(b) Science

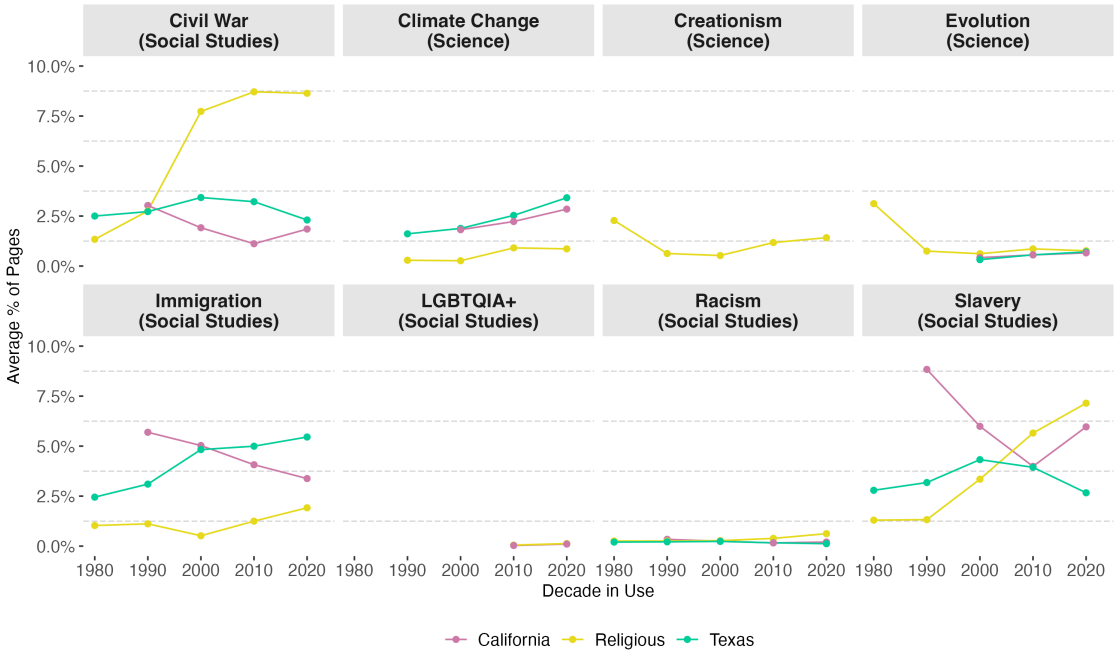


(c) Social Studies



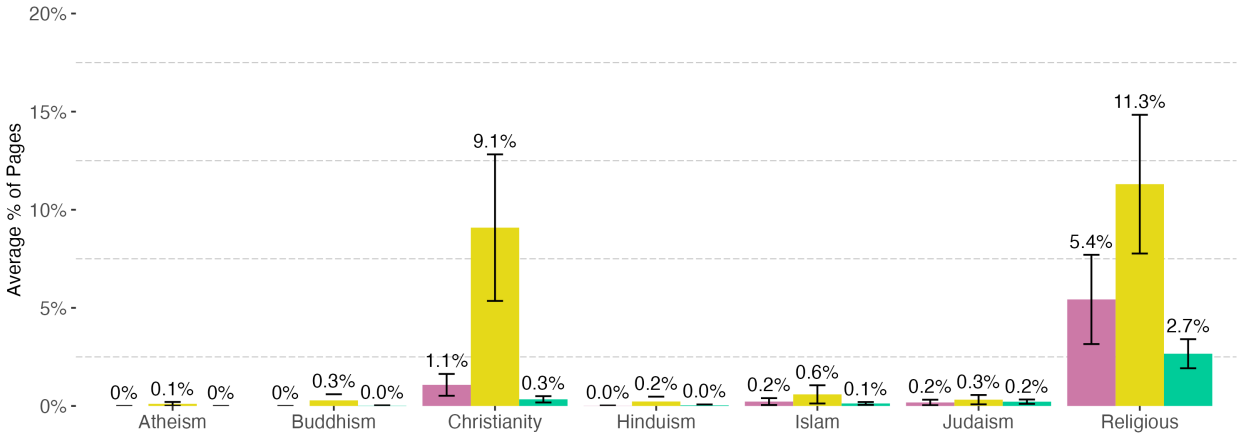
Notes: This figure shows the top topics (e.g. topics that appear on the highest average percentage of pages) within a collection. To calculate the average percentage of pages containing each topic, we first calculate the percentage of pages for each collection, subject, grade, and year in use. We then average these percentages at the collection level, to ensure that pages in each grade, subject, and year in use are weighted equally in the final averages.

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



Notes: This figure illustrates the average percentage of pages within a specific subject as denoted in the parentheses that contain a specific topic over time. To calculate the average percentage of pages containing each topic, we first calculate the percentage of pages for each collection, grade, and year in use. We then average these percentages at the collection level and decade level, to ensure that pages in each grade and year in use are weighted equally in the subject and collection specific final averages.

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



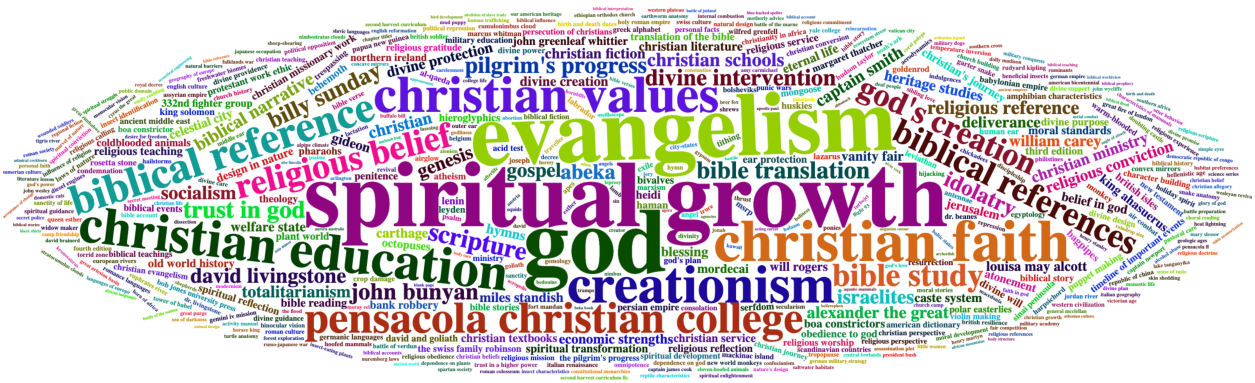
Notes: This figure illustrates the average percentage of pages within Social Studies textbooks that contain a religious topic or topics related to a specific religion. To calculate the average percentage of pages containing each topic, we first calculate the percentage of pages for each collection, grade, and year in use. We then average these percentages at the collection level, to ensure that pages in each grade and year in use are weighted equally in the final averages.

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



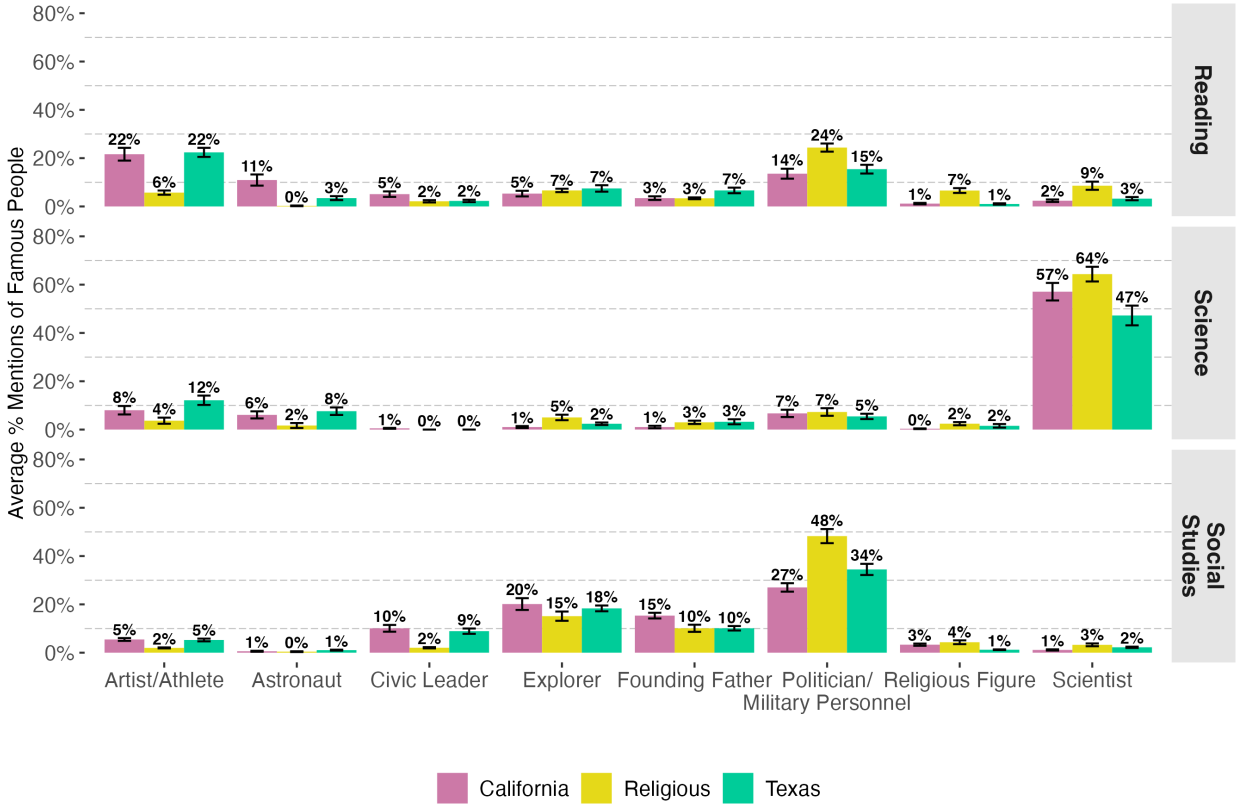
(c) Topics and Values Exclusive to Texas Collection



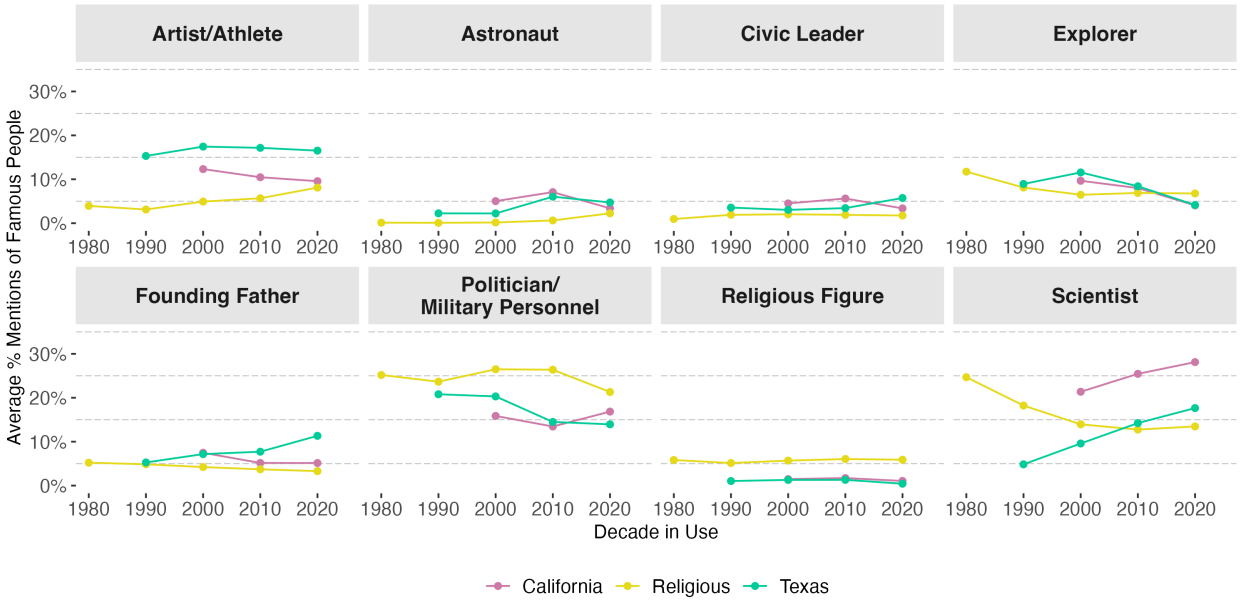
Notes: Each word cloud displays the topics and values metadata tags that appear exclusively within a single collection and not in either of the other two. Word size is proportional to frequency of occurrence across items in that collection. Publisher-specific tags (e.g., A Beka Book, BJU Press, Discovery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

Figure A.9: Historical Figures, by Occupation

(a) By Subject



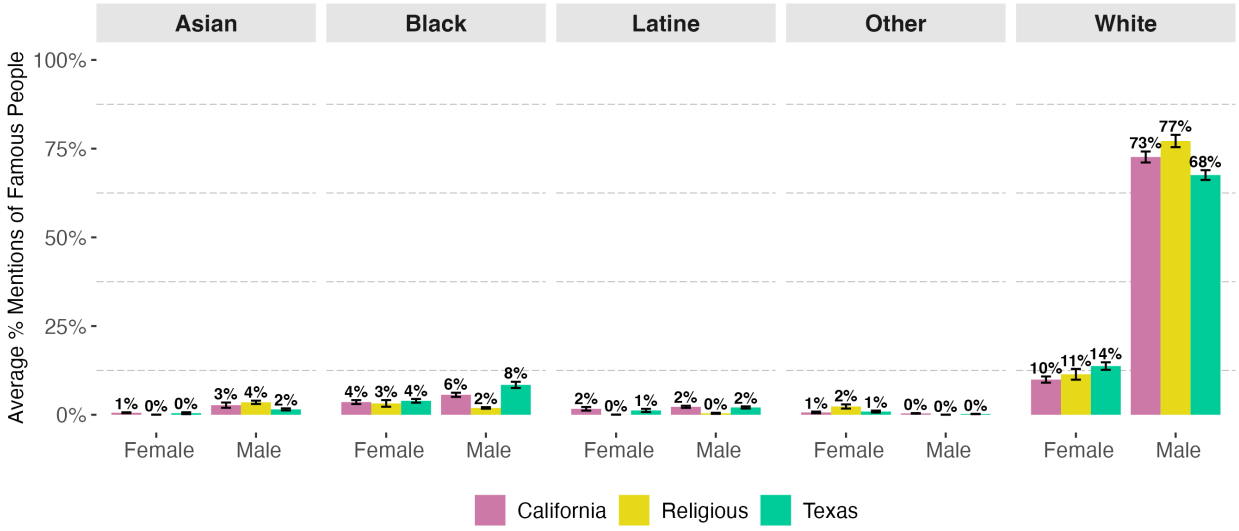
(b) Over Time



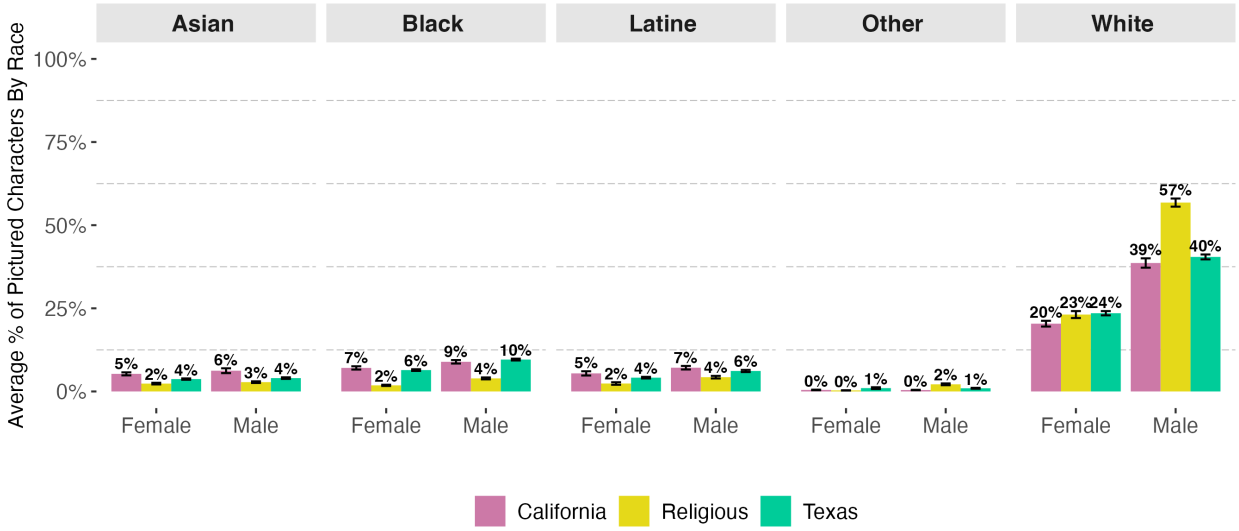
Notes: This figure shows the average percentage of mentions of famous people by their occupation by subject (Panel (a)) and over time (Panel (b)). To determine the average percentage of mentions for each occupation, we first calculate the percentage of mentions across each subject, grade, and year in use. We then average these percentages at the collection-by-subject or collection level to ensure that pages from each grade, subject, and year in use are weighted equally in the final averages.

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

(a) Text (Famous Individuals)



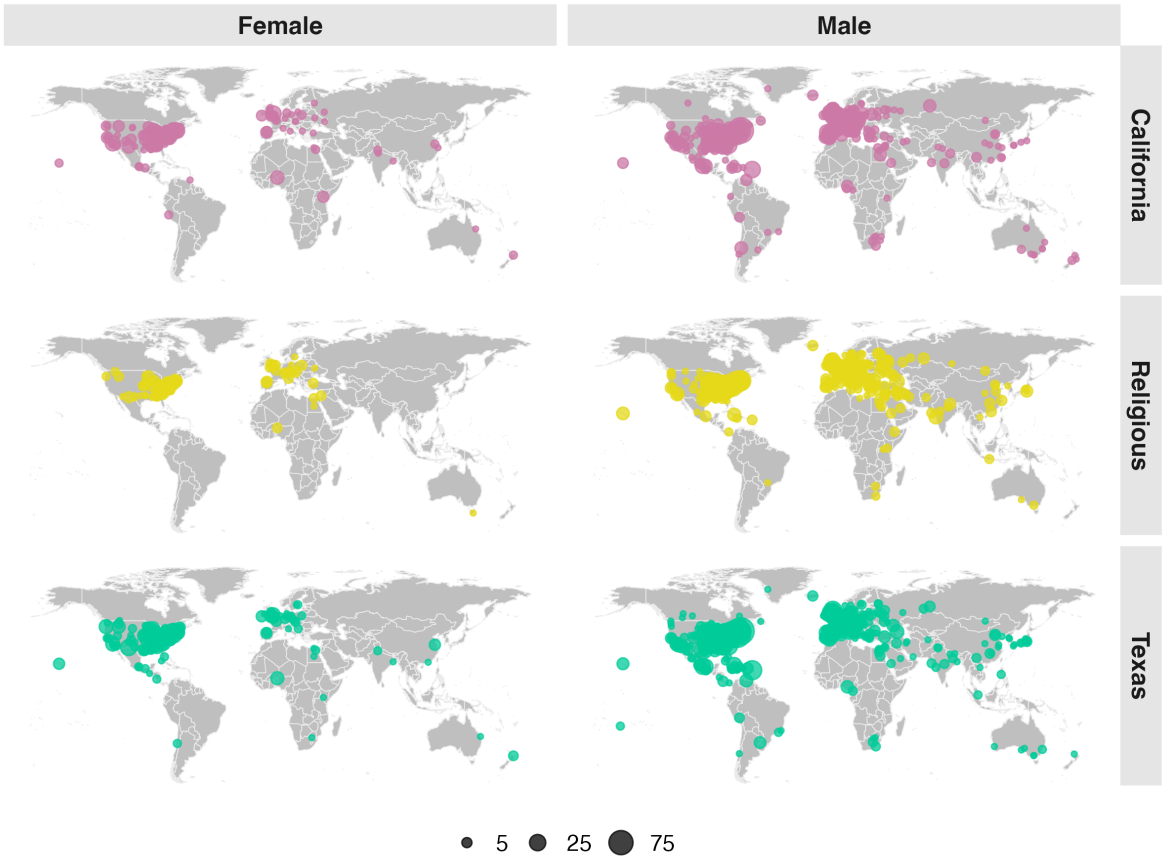
(b) Images



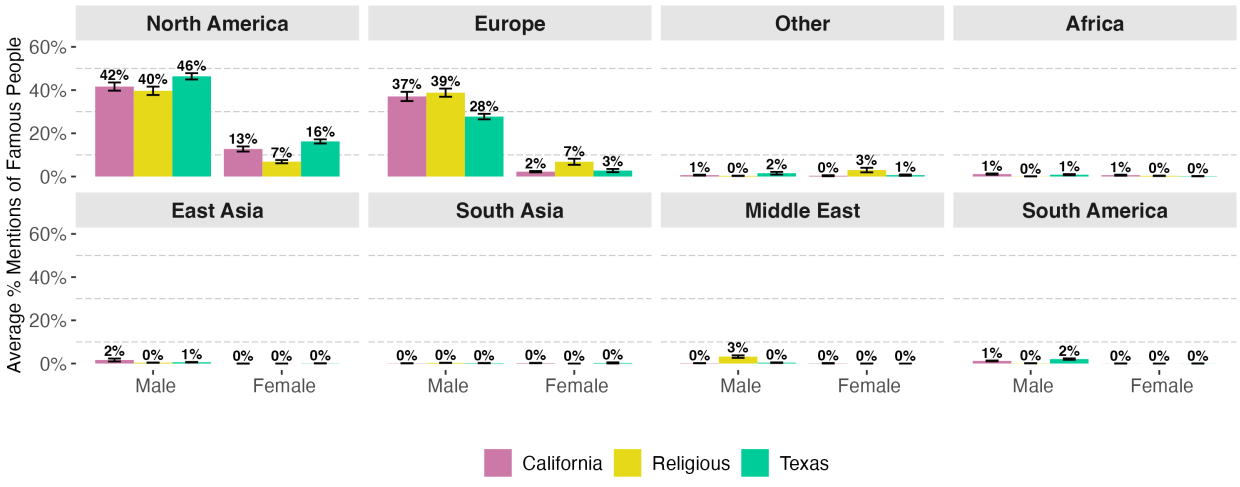
Notes: Panel (a) shows the average percentage of famous people that identify as a specific race and gender out of all mentions of famous people by collection. To determine the average percentage of mentions by collection, we first calculate the percentage of mentions of famous people of a specific gender by race across each subject, grade, and year in use. We then average these percentages at the decade level to ensure that pages from each grade, subject, and year in use are weighted equally in the final averages. Panel (b) shows the average percentage of faces classified as a specific race and gender by our feature classification model described in Section V.A.2 by collection and subject. To determine the average percentage, we first calculate the percentage of faces by race within each collection, subject, grade, and year in use. We then average these percentages at the collection level to ensure that pages from each grade, subject, and year in use are weighted equally.

Figure A.11: Birthplaces of Historical Figures by Gender

(a) Map



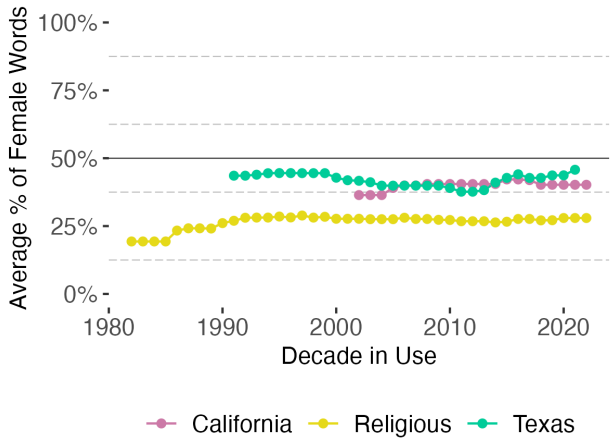
(b) Percent of Mentions



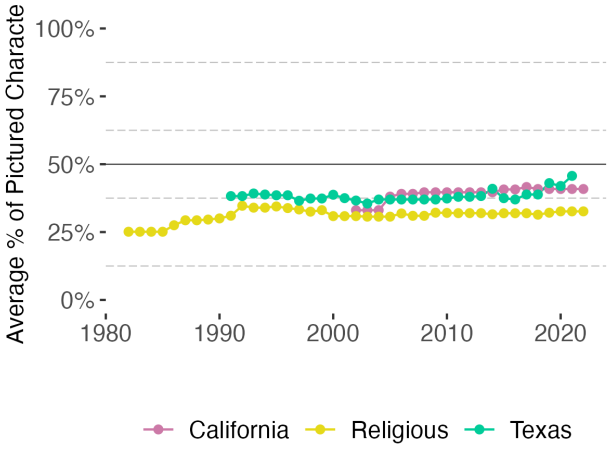
Notes: This figure shows the birthplaces of all famous people mentioned in a collection by gender. In Panel (a), the size of the point represents the number of mentions. In Panel (b) we show the average percentage of mentions for each birthplace by gender.

Figure A.12: Female Representation

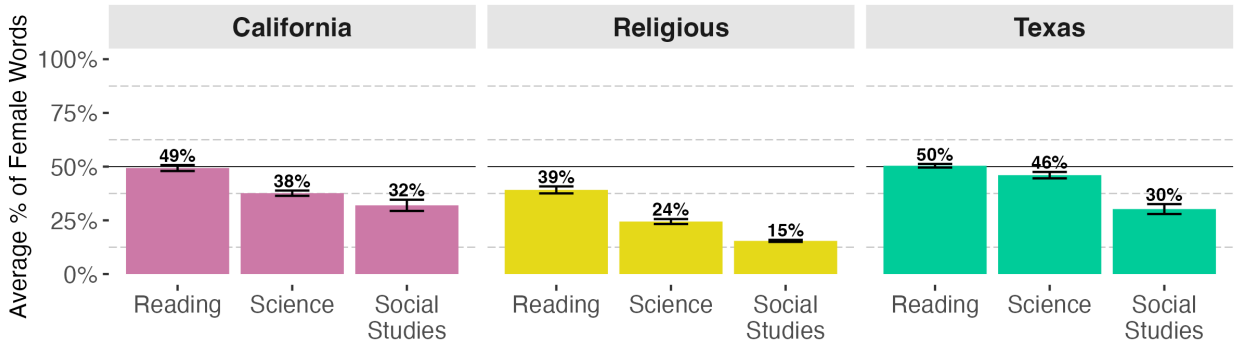
(a) Text Over Time



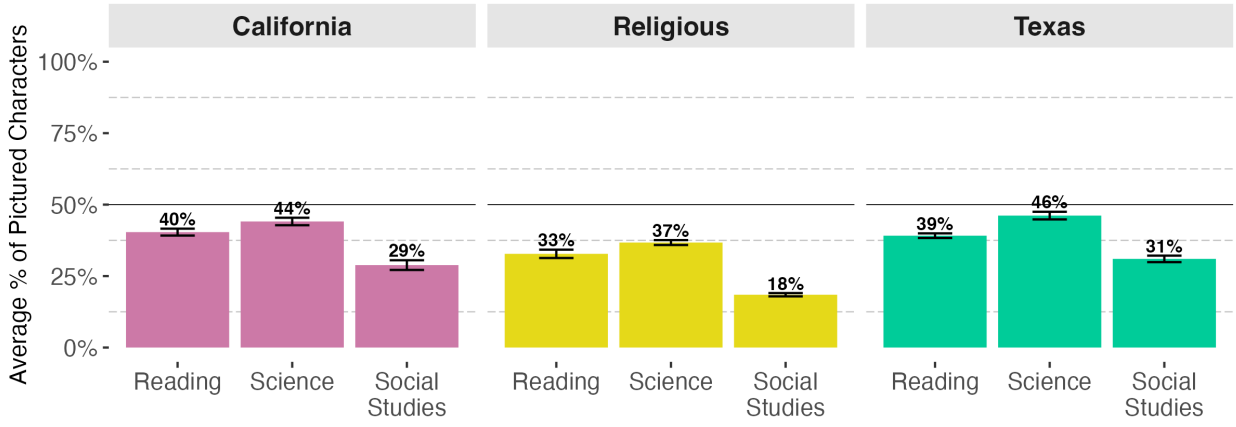
(b) Images Over Time



(c) Text by Subject



(d) Images by Subject



Notes:

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 V. Standard errors clustered at the textbook level. *p<0.1; **p<0.05; ***p<0.01

Table B.4: 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.

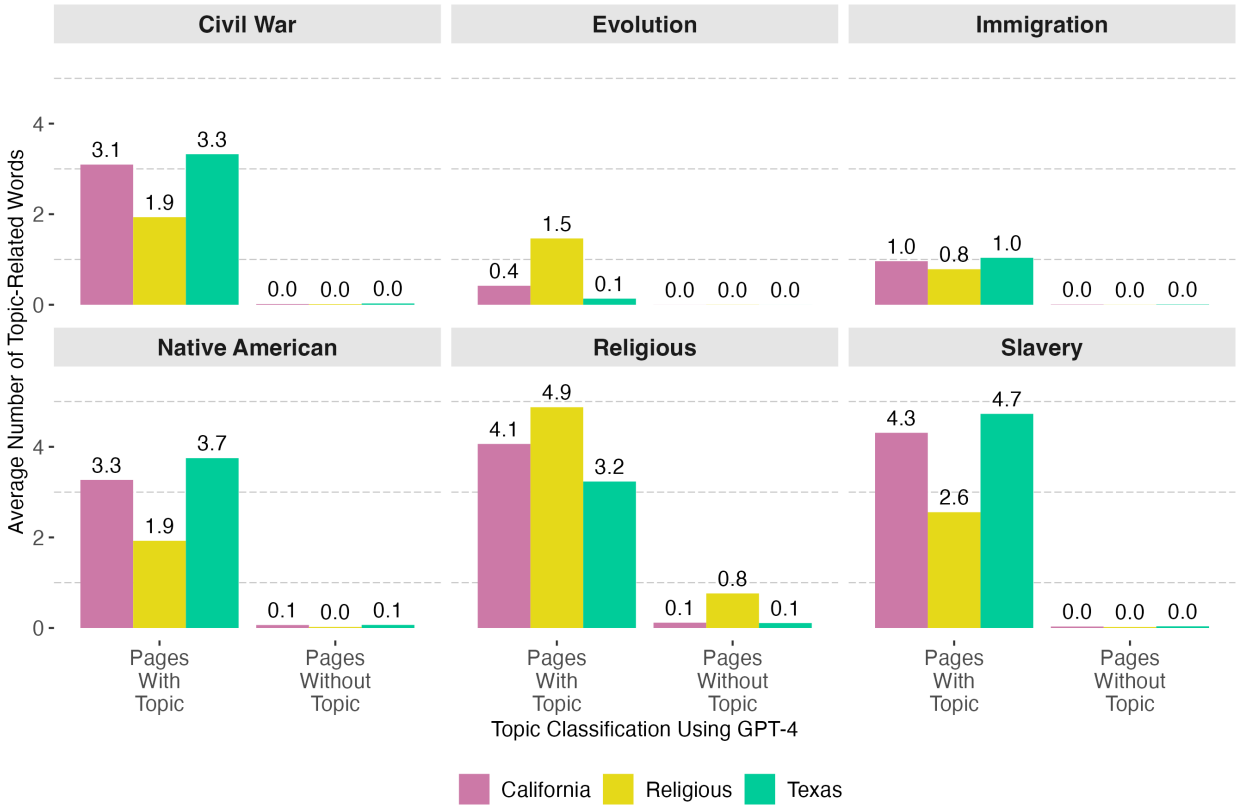
Note that for the gemini model this is only the accuracy of characters in our sample that **also** had a face detected and not the overall accuracy of the model. We merged characters detected and faces detecting using the largest overlap of bounding boxes. We were unable to match all of detected faces 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 is find 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 yield 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.[15]

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.

[15]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`) with temperature set to 0.0.

In step one, we identify pages containing visual content and separate them from pages containing only text. We then filter out pages containing only text before moving to the second stage to minimize computational expenses by including pages devoid of illustrations. For each scanned page, the raw image bytes are submitted to the LMM 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 Gemini model to output a JSON object capturing: (i) scene-level data (what is happening or setting description) and (ii) character-level data (bounding-box coordinates, race, gender, human or non-human, actions, location). Note that we recode the model's race output into categories (Black, Latine, Asian, White, Indigenous, Unsure, Other, Not applicable).

D.B 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, which allows a 16,000 token limit for total input and response. The parameter *temperature* defines the freedom or creativity we allow the model to have for its responses, and is set to 1. The optimal value for the temperature parameter will usually depend on the type of task we ask GPT to do, and to the best of our knowledge there is not a benchmark or standard for this. While higher temperature values give the model more room to hallucinate (create fake or made-up responses), lower ones make responses more deterministic, potentially limiting its flexibility to provide reasonable responses. 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.C Skin Color

To extract and classify skin color, we begin by identifying the facial area of interest in the pictured face by "segmenting" the skin. To do this, we use a face landmarker model,[16] 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"). Then, we extract the dominant colors in the segmented skin by partitioning the pixels of the segmented image into k-means clustering. This groups each pixel into k clusters that share the most similar RGB color space, and each

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

cluster is represented by the RGB value of its center. Finally, we collapse the remaining segmented skin into a single "representative color" by taking a weighted average of the dominant skin colors (where the weights are determined by the size of the cluster from which each dominant color was drawn). To construct a continuous measure of how dark or light the skin color of a detected face is, we convert the center of each color cluster in RGB space into the perceptually linear CIE L*a*b* color space and take their weighted average to construct a representative skin color of the face. We use the L* value of this representative skin color which ranges from 0 (darkest) to 100 (lightest) as our continuous measure of skin tint that serves as our measure of skin color.

D.D Top Relative Words.

To characterize the contexts most strongly associated with female versus male pronoun, we identify "top relative words" using the à la carte pronoun-instance embeddings or image description embeddings for pictured characters described in Section V. This approach follows a similar methodology outlined in Manzini et al. (2019) and Charlesworth et al. (2022), where the "top relative words" are found by computing the group-to-word association between a single word and the target (or comparison) group. However, in our implementation the analysis is conducted at the pronoun or pictured character instance level rather than using only one average embedding per group.

Let v_{ic}^{g} denote the embedding for pronoun or pictured character 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 pronoun embeddings within collection c. For a female pronoun instance in collection c, we define the relative query vector as:

$$q_{ic}^{f} = v_{ic}^{f} - \bar{v}_c^m, \tag{D.2}$$

and analogously, for a male pronoun instance:

$$q_{ic}^{m} = v_{ic}^{m} - \bar{v}_c^f, \tag{D.3}$$

Intuitively, these relative query vectors capture the portion of a pronoun's or image description's contextual embedding that is distinctive relative to the opposite gender within the same collection.

For each pronoun instance, we then 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.

We estimate whether these average VAD scores differ by gender within collections using regressions of the form

$$Y_{ib}^{k} = \alpha + \sum_{c} \lambda_c \mathbf{1}\{C_i = c\} + \sum_{c} \beta_c \left(\mathbf{1}\{C_i = c\} \times Female_i\right) + X_i'\gamma + \varepsilon_{ib}, \tag{D.4}$$

where Y_{ib}^{k} is the mean score for dimension $k \in \{V, A, D\}$ for pronoun or image description

embedding i in book b, C_i denotes the textbook collection, and X_i includes controls for subject, grade, and year in use. Standard errors are clustered at the book level. Under this specification, β_c captures the female-minus-male difference in the average VAD score of top relative words within collection c which we refer to as "gender-skew" of a VAD dimension in a collection. Since males are the excluded group, if β_c is positive, this means the VAD dimension is more female skewed in that collection (e.g. females are talked about in more positive contexts than males if β_c is positive when the dependent variables is Y_{ib}^V).

D.E 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 pronoun embedding or image description 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 pronoun or image description embedding i, we measure association with domain D as the average cosine similarity between the embedding and the words in that domain:

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

This quantity can be interpreted as the average directional similarity between the gendered context in text or images and the semantic domain. We then estimate collection-specific gender differences in domain association using

$$s_{iD} = \alpha_D + \sum_c \lambda_{Dc} \mathbf{1}\{C_i = c\} + \sum_c \beta_{Dc} \left(\mathbf{1}\{C_i = c\} \times Female_i\right) + X_i' \gamma_D + \varepsilon_{iDb}, \tag{D.6}$$

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. We called this the "gender-skew" of a domain.

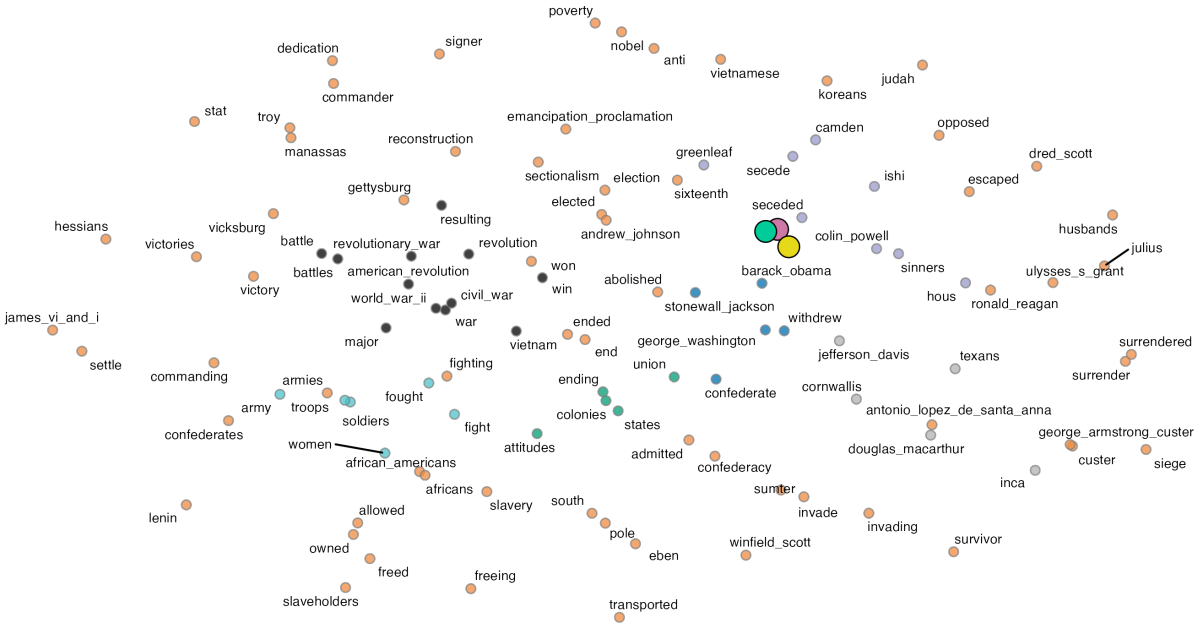
D.F Embedding Regressions

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 seman-

tic 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.

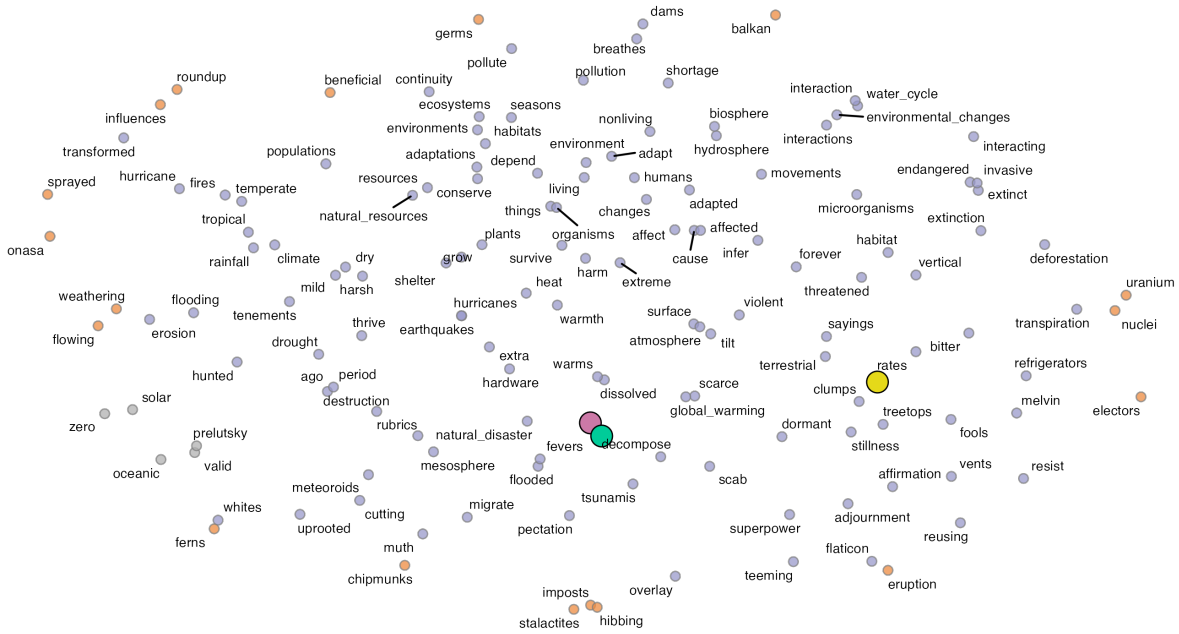
E Topic Embeddings

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



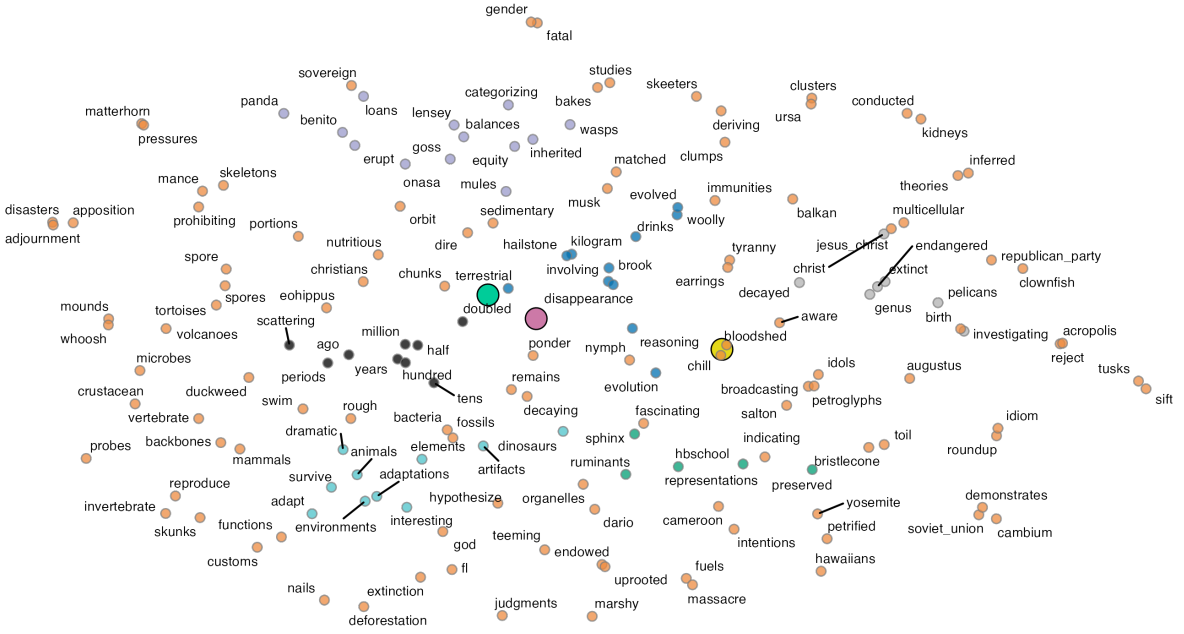
Notes: This figure shows the cosine similarity between a given topic vector and a word vector for the unique union of the top 10 words associated with each collection's topic vector.

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



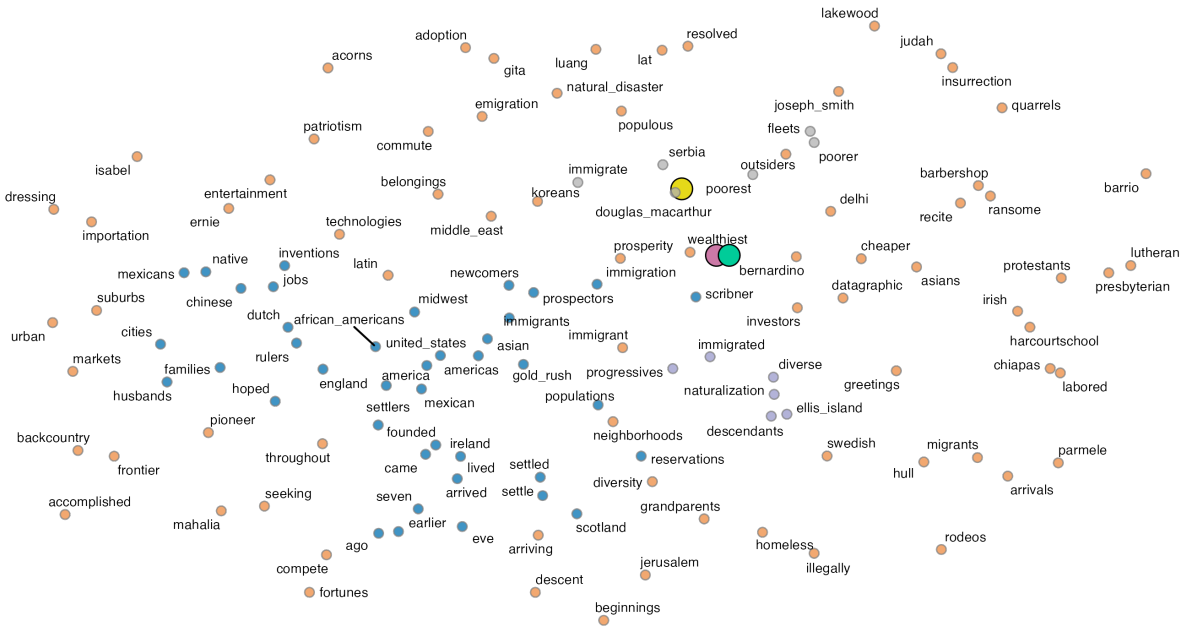
Notes: This figure shows the cosine similarity between a given topic vector and a word vector for the unique union of the top 10 words associated with each collection's topic vector.

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



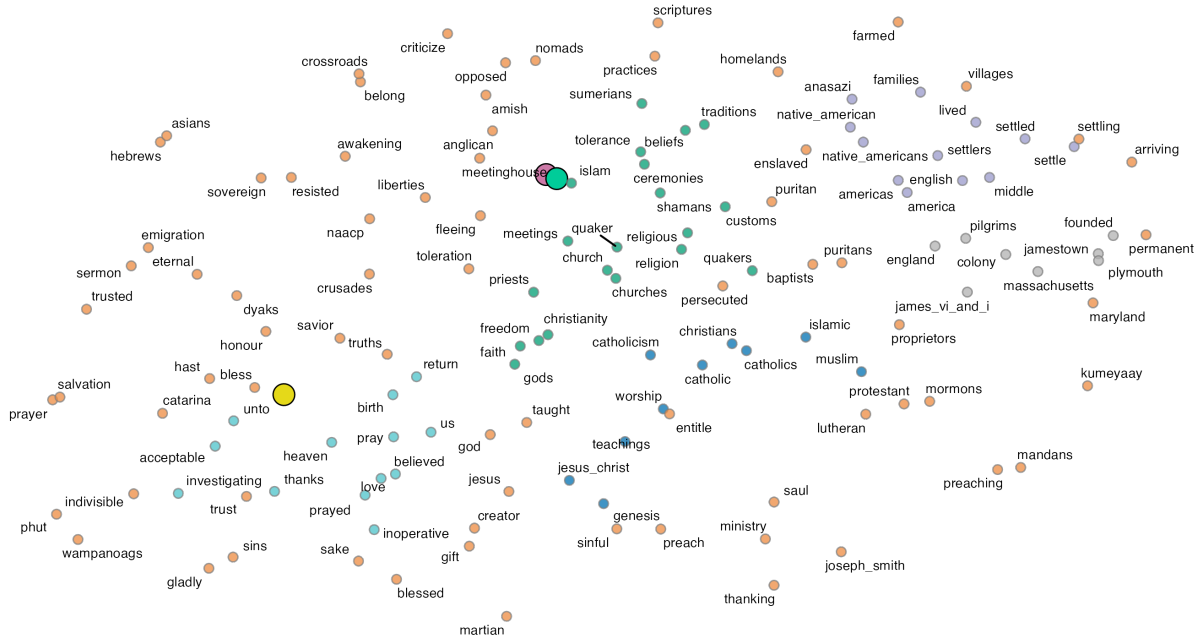
Notes: This figure shows the cosine similarity between a given topic vector and a word vector for the unique union of the top 10 words associated with each collection's topic vector.

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



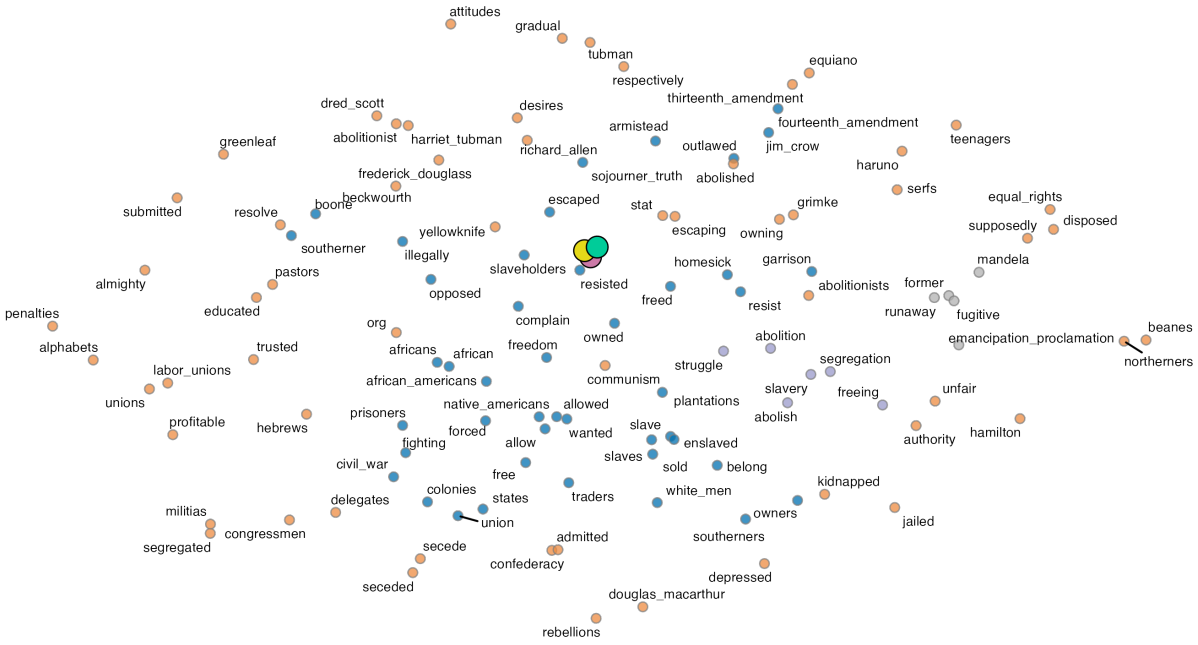
Notes: This figure shows the cosine similarity between a given topic vector and a word vector for the unique union of the top 10 words associated with each collection's topic vector.

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



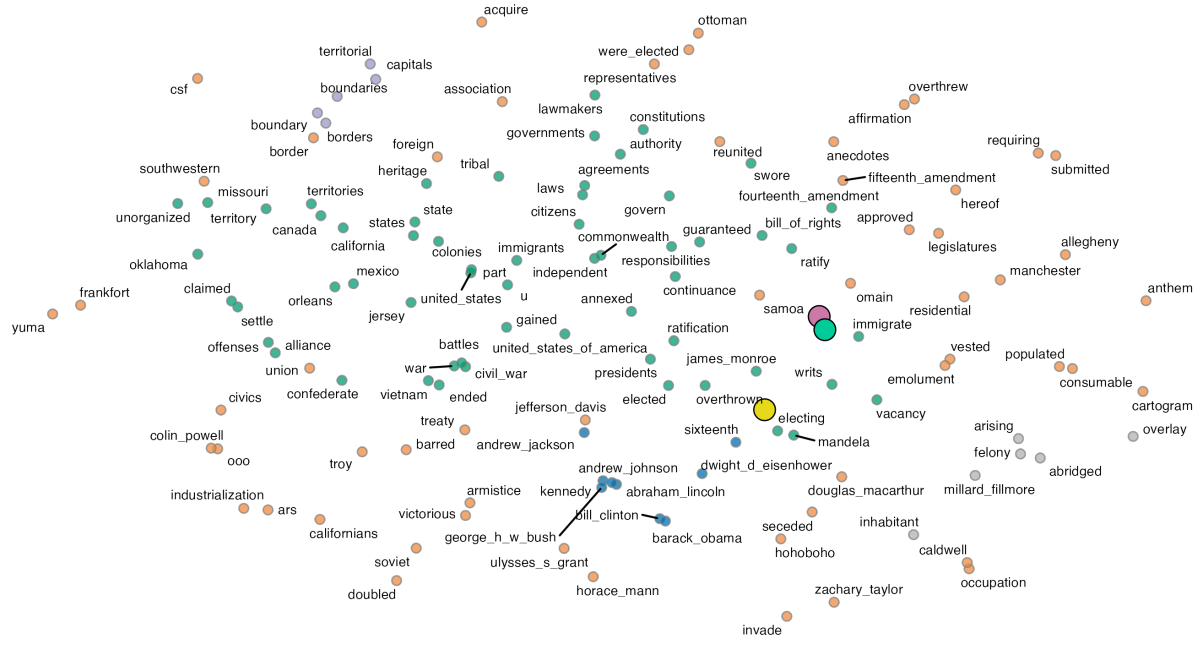
Notes: This figure shows the cosine similarity between a given topic vector and a word vector for the unique union of the top 10 words associated with each collection's topic vector.

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



Notes: This figure shows the cosine similarity between a given topic vector and a word vector for the unique union of the top 10 words associated with each collection's topic vector.

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



Notes: This figure shows the cosine similarity between a given topic vector and a word vector for the unique union of the top 10 words associated with each collection's topic vector.